

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)特許出願公表番号

特表2002-513487

(P2002-513487A)

(43)公表日 平成14年5月8日(2002.5.8)

(51)Int.Cl. <sup>7</sup>	識別記号	FI	キーワード(参考)
G06F 17/30	170	G06F 17/30	170D
H04N 7/173	640	H04N 7/173	640A

審査請求 未請求 予備審査請求 有 (全50頁)

(21)出願番号 特願平10-548376  
(86)(22)出願日 平成10年5月5日(1998.5.5)  
(85)翻訳文提出日 平成11年11月5日(1999.11.5)  
(86)国際出願番号 PCT/US98/09124  
(87)国際公開番号 WO98/50869  
(87)国際公開日 平成10年11月12日(1998.11.12)  
(31)優先権主張番号 60/045,637  
(32)優先日 平成9年5月5日(1997.5.5)  
(33)優先権主張国 米国(US)  
(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), CA, JP, KR, US

(71)出願人 ザ トラスティーズ オブ コロンビア  
ユニヴァーシティ イン ザ シティ オ  
ブ ニューヨーク  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク ブロードウェイ アンド  
ワンハンドレッドシックスティーン  
ストリート(番地なし)  
(72)発明者 チャン シーファー  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク リヴァーサイド ドライ  
ヴ 560 アパートメント18ケイ  
(74)代理人 弁理士 杉村 曉秀 (外2名)

最終頁に続く

(54)【発明の名称】 オブジェクト指向コンテンツに基づくビデオ・サーチ用のアルゴリズム及びシステム

(57)【要約】

相互作用的なネットワークを介して1つ或はそれ以上ビデオ・クリップから1つ或はそれ以上のビデオ・オブジェクトを探し出させることをユーザに許容するオブジェクト指向方法及びシステムが開示されている。本システムは、ビデオ・クリップ用の記憶装置(111)とビデオ・オブジェクト属性のデータベースとを含む1つ或はそれ以上のサーバ・コンピュータ(110)、通信ネットワーク(120)、並びに、クライアント・コンピュータ(130)を備えている。クライアント・コンピュータは、運動軌跡情報を含んでいるビデオ・オブジェクト属性情報を指定する照会インタフェースと、前記サーバ・コンピュータ内に保存されたビデオ・オブジェクト属性をブラウズするブラウザと、相互作用的なビデオ・プレイヤーとを備えている。

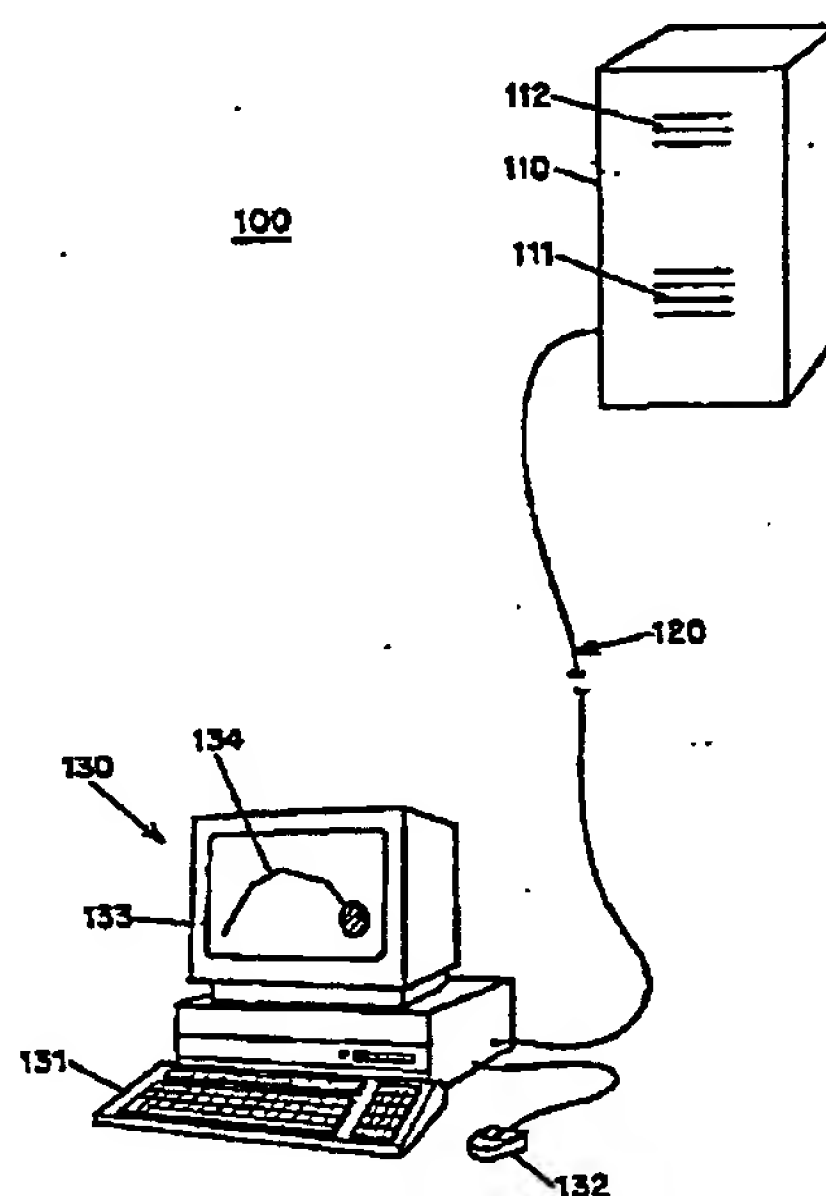


FIG. 1

BEST AVAILABLE COPY

## 【特許請求の範囲】

1. 相互作用的なネットワークを介して1つ或はそれ以上ビデオ・クリップから1つ或はそれ以上のビデオ・オブジェクトを探し出させることをユーザに許容するオブジェクト指向システムであって、
  - a. 前記1つ或はそれ以上のビデオ・クリップのための記憶装置と、前記ビデオ・クリップに対応するビデオ・オブジェクト属性の1つ或はそれ以上のデータベースのための記憶装置とを備える1つ或はそれ以上のサーバ・コンピュータと、
  - b. 前記1つ或はそれ以上のサーバ・コンピュータに結合された通信ネットワークであり、前記サーバ・コンピュータから前記1つ或はそれ以上のビデオ・クリップの伝送を許容する通信ネットワークと、
  - c. 前記通信ネットワークに結合されたクライアント・コンピュータであり、
    - i. 運動軌跡情報を含むビデオ・オブジェクト属性情報を指定する照会インタフェースと、
    - ii. 前記照会インタフェースに結合されて、前記選択されたビデオ・オブジェクト属性情報を受け取るブラウザ・インタフェースであり、前記通信ネットワークによって前記サーバ・コンピュータ内の前記記憶されたビデオ・オブジェクト属性をブラウズして、前記指定されたビデオ・オブジェクト属性に最も符合する属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定するブラウザ・インタフェースと、
    - iii. 前記サーバ・コンピュータから、前記決定された1つ或はそれ以上のビデオ・オブジェクトに対応するビデオ・データから成る1つ或はそれ以上の伝送された連続フレームを受け取る相互作用的なビデオ・プレイヤーとを有するクライアント・コンピュータとを具備するシステム。
2. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、運動軌跡データベースを含む、請求項1に記載のシステム。
3. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベ-

スの内の1つが、時空間データベースを含む、請求項1に記載のシステム。

4. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、形状データベースを含む、請求項1に記載のシステム。
5. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、色彩データベースを含む、請求項1に記載のシステム。
6. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、質感データベースを含む、請求項1に記載のシステム。
7. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、パン撮りデータベースを含む、請求項1に記載のシステム。
8. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、ズーム撮りデータベースを含む、請求項1に記載のシステム。
9. ビデオ・データから成る前記1つ或はそれ以上の連続フレームが圧縮フォーマットで前記サーバ・コンピュータに保存されている、請求項1に記載のシステム。
10. 前記1つ或はそれ以上の指定されたビデオ・オブジェクト属性の各々を前記サーバ・コンピュータ内に対応して保存されたビデオ・オブジェクト属性と比較して、各ビデオ・オブジェクト属性に1つずつの候補ビデオ・シーケンスのリストを生成する手段を前記サーバ・コンピュータの内の少なくとも1つ或はそれ以上の内部に更に含む、請求項1に記載のシステム。
11. 前記比較手段に結合されて前記候補リストを受け取る手段であり、前記候補リストに基づいて、前記選択されたビデオ・オブジェクト属性に最も符合する集合的な属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定する手段を前記サーバ・コンピュータ内に更に含む、請求項10に記載のシステム。
12. 前記照会されたビデオ・オブジェクト属性情報が、2つ以上のビデオ・オブジェクトに対する属性を含み、前記比較手段が、各ビデオ・オブジェクトに対する前記1つ或はそれ以上の指定されたビデオ・オブジェクト属性の各々を

前記サーバ・コンピュータ内の対応する保存されたビデオ・オブジェクト属性と比較して、各ビデオ・オブジェクトに対する各ビデオ・オブジェクト属性に1つずつの、候補ビデオ・シーケンスのリストを生成し、前記決定手段が、各照会されたビデオ・オブジェクトに対する前記候補リストに基づいて、前記選

択されたビデオ・オブジェクト属性に最も符合する集合的属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定している、請求項11に記載のシステム。

13. 少なくとも1つの認識可能な属性を含むビデオ・クリップからビデオ・オブジェクトを抽出する方法であって、

a. ビデオ・データで表示される前記少なくとも1つの属性の異なる変体に値を決定すると共に該値を割り当てることによって、前記ビデオ・データから成る現行フレームを量子化することによって、量子化フレーム情報を生成する工程と、

b. 前記少なくとも1つの属性に基づいて、前記ビデオ・データから成るフレーム上のエッジ検出を実行して、該フレーム内のエッジ点を決定することによって、エッジ情報を生成する工程と、

c. 先行するフレームから1つ或はそれ以上に区分された領域を規定する情報を受け取る工程と、

d. 前記現行フレームから、前記受け取られた区分領域を前記量子化フレーム情報及び前記生成エッジ情報と比較することによって、前記少なくとも1つの属性を共有するビデオ情報から成る領域を抽出する工程との諸工程を含む方法。

14. 前記属性が色彩であり、量子化工程が、前記現行フレームを均一色彩スペース情報に変換し、該色彩スペース情報を適応的に量子化し、前記パレットを濾過してそこからノイズを排除することを含む、請求項13に記載の方法。

15. 前記適応的な量子化工程が、クラスタリング・アルゴリズムでの量子化を含む、請求項14の方法。

16. 前記エッジ検出工程が、前記現行フレームにCannyエッジ検出を適用して

、前記エッジ情報をエッジ・マップとして生成することを含む、請求項13に記載の方法。

17. 前記抽出工程が、

a. 前記受取領域の内の1つを現行の量子化され且つエッジ検出されたフレームに投影することによって、ビデオ・データから成る前記現行フレーム内

の領域を抽出すべくインターフレーム投影を実行して、前記領域の任意の移動を時間的に追跡することと、

b. 前記現行フレーム内の隣接する抽出領域を併合すべくイントラフレームを実行することと、

を含む、請求項13に記載の方法。

18. 前記属性が色彩であり、前記インターフレーム投影工程が、

a. 前記先行フレームからの前記受取領域を前記現行フレーム内へ投影して、領域を時間的に追跡する工程と、

b. 前記受取領域と一貫性を有するか或は新規領域として、前記現行フレーム内における各非エッジ・ピクセルをラベル表示する工程と、

c. 隣接する領域をリンクすべく、前記ラベルから接続グラフを生成する工程と、

の諸工程を含むことから成る、請求項17に記載の方法。

19. 前記イントラフレーム区分工程が、

a. 所定閾値よりも小さい色彩距離を有する全ての隣接領域を新規領域として併合する工程と、

b. 前記新規領域に対する平均色彩を決定する工程と、

c. 前記接続グラフを更新する工程と、

d. 前記新規領域に、前記併合領域に先行して割り当てられたラベルから新規ラベルを割り当てる工程と、

e. 前記併合領域を排除する工程と、

の諸工程を含むことから成る、請求項18に記載の方法。

20. 前記抽出工程が、隣接領域に為す前記イントラフレーム区分の後に残存す

る前記現行フレーム内の全てのエッジをラベル表示して、それらの各ラベル表示されたエッジが前記現行フレーム内のビデオ・オブジェクトの境界を画成するように為す工程を更に含むことから成る、請求項17に記載の方法。

21. 前記抽出工程が、所定閾値以下のサイズを有する領域を削除することによって、前記抽出領域を簡略化する工程を更に含むことから成る、請求項20に記載の方法。

22. e. ビデオ情報から成る将来フレームを受け取る工程と、

f. 前記現行フレーム内のビデオ情報から成るブロックと、前記将来フレーム内のビデオ情報から成るブロックとの間の階層的なブロック突き合わせを実行することによってビデオ情報から成る前記現行フレームの光学的フローを決定する工程と、

g. 前記光学的フローに基づいて、ビデオ情報から成る前記抽出領域上の運動推定を実行する工程と、

の諸工程を更に含む、請求項13に記載の方法。

23. 前記現行フレーム内の決定領域をサイズ及び持続時間でグループ分けする工程を更に含む、請求項22に記載の方法。

24. 前記現行フレーム内の決定領域をその内部の移動オブジェクトを決定することでグループ分けする工程を更に含む、請求項22に記載の方法。

25. それぞれが所定軌跡で時間的に移動する1つ或はそれ以上のビデオ・オブジェクトを含んでいる1つ或はそれ以上のビデオ・クリップから、ユーザ入力調査照会に最も符合するビデオ・クリップを探し出す方法であって、

a. 少なくとも1つのビデオ・オブジェクト軌跡を画成する調査照会を受け取る工程と、

b. 前記受取照会と、1つ或はそれ以上の予め画成されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の合計距離を決定する工程と、

c. 前記受取照会から最小距離を有する前記画成されたビデオ・オブジェクト軌跡の内の1つ或はそれ以上を選択することによって、前記最も符合したビデオ・クリップ或は複数のビデオ・クリップを探し出す工程と、

の諸工程を含む方法。

26. 前記保存されたビデオ・クリップが正規化されて、再生フレーム率が所定時間スケールまで増倍されるように為され、前記受取照会を各正規化ビデオ・クリップへマッピングして、そのマッピングされた受取照会を前記正規化ビデオ・クリップによって規定された各ビデオ・オブジェクトまで増倍することによって前記受取調査照会を正規化する工程を更に含み、前記決定工程が前記正規化された受取照会と前記正規化されたビデオ・オブジェクト軌跡との

間の合計距離を決定してることから成る、請求項25に記載の方法。

27. 前記決定工程が、前記受取ビデオ・オブジェクト軌跡と前記1つ或はそれ以上の予め規定されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の空間的距離の比較を為すことを含むことから成る、請求項25に記載の方法。

28. 前記決定工程が、前記受取ビデオ・オブジェクト軌跡と前記1つ或はそれ以上の予め規定されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の時空間的距離の比較を為すことを含むことから成る、請求項25に記載の方法。

29. それぞれが所定属性を有する1つ或はそれ以上のビデオ・オブジェクトを含んでいる1つ或はそれ以上のビデオ・クリップから、ユーザ入力調査照会に最も符合するビデオ・クリップを探し出す方法であって、

a. 前記ビデオ・クリップ内の1つ或はそれ以上の異なるビデオ・オブジェクトに対する1つ或はそれ以上の属性を規定する調査照会を受け取る工程と、

b. 前記ビデオ・クリップを調査して、前記規定された属性の内の少なくとも1つを所定閾値に符合させる1つ或はそれ以上のビデオ・オブジェクトを探し出す工程と、

c. 前記探し出されたビデオ・オブジェクトから、前記1つ或はそれ以上の異なるビデオ・オブジェクトを含む1つ或はそれ以上のビデオ・クリップを決定する工程と、

d. 前記調査照会によって規定された前記1つ或はそれ以上のビデオ・オブジェクトと前記探し出されたビデオ・オブジェクトとの間の距離を計算するこ

とによって、前記決定されたビデオ・クリップから最も符合するビデオ・クリップを決定する工程と、  
の諸工程を含む方法。

30. 前記1つ或はそれ以上の属性が色彩を含み、前記符合工程が、前記照会されたビデオ・オブジェクトの各々に対する平均色彩を決定して、該平均色彩をデータベースに保存された色彩情報と比較することを含むことから成る、請求項29に記載の方法。

31. 前記1つ或はそれ以上の属性が質感を含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対する粗さ、コントラスト、並びに、配向を決定して、該粗さ、コントラスト、並びに、配向をデータベース内に保存された粗さ、コントラスト、並びに、配向の情報と比較することを含むことから成る、請求項29に記載の方法。

32. 前記1つ或はそれ以上の属性が形状を含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対する主要軸線に沿っての固有値を決定して、該固有値をデータベース内に保存された形状情報と比較することを含むことから成る、請求項29に記載の方法。

33. 前記1つ或はそれ以上の属性がサイズを含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対するパーセンテージ面積を決定して、該面積をデータベース内に保存された面積情報と比較することを含むことから成る、請求項29に記載の方法。

34. 前記ビデオ・クリップが関連されたテキスト情報を含み、前記参照照会が前記1つ或はそれ以上の異なるビデオ・オブジェクトに対応しているテキスト特性の定義を更に含み、前記関連されたテキスト情報を調査して前記テキスト特性に最も符合するテキストを探し出す工程を更に含む、請求項29に記載の方法。

35. 前記最も符合するビデオ・クリップが前記決定ビデオ・クリップ及び前記探し出されたテキストから決定されることから成る、請求項30に記載の方法。  
。

## 【発明の詳細な説明】

オブジェクト指向コンテンツに基づくビデオ・サーチ用のアルゴリズム及びシステム関連出願の相互参照

この出願は、1997年5月5日出願の合衆国予備出願第60/045,637号に関連されて、そこから優先権が主張されている。

発明の背景

## 1. 発明の分野

本発明はビジュアル（可視）情報をサーチ（調査）して検索する技術に関し、より詳細には移動しているビジュアル情報を探し求めて検索すべくコンテンツに基づく調査照会の使用に関する。

## 2. 関連技術の説明

過去数年間で、インターネットは円熟期に達して、マルチメディア応用は広範な用途にわたってきており、容易に利用可能なデジタル・ビデオ情報のストックは絶えず増え続けてきている。帯域幅要件を管理可能なレベルまで低減するために、そうしたビデオ情報は、例えばJ P E G、MotionJ P E G、M P E G-1、M P E G-2、M P E G-4、H. 261、或は、H. 263等の標準フォーマットである圧縮ビットストリームの形態であるデジタル環境内に一般的には保存又は記憶されている。現在、海及び山からスキー及び野球まであらゆるものを表す数百万もの異なる静止画像及び動画画像がインターネットを通じて利用可能となっている。

デジタル・フォーマットで利用可能な増大する豊富なビデオ情報に伴って、そうした情報を有意義に構成しサーチし尽くす必要性が差し迫ってきている。詳細にはユーザが定義した照会に応じて、保存されたビデオ情報内に埋め込まれたビデオ・オブジェクトの形状或は運動の特性等、自由裁量の所定規準に合致するビデオ情報の内の特定の片々をサーチして検索できるビデオ・サーチエンジンに基づくコンテンツをユーザは益々要望している。

この必要性に応じて、ビデオのサーチ及び検索のアプリケーションを開発する

幾つかの試みが為されてきた。現存する技術は2つの異なるカテゴリー：例示による照会（又は問い合わせ）（“QBE”）及びビジュアル・スケッチング(visual sketching)に入る。

画像検索に関連して、QBEシステムの例は、QBIC、PhotoBook、VisualSEEK、Virage、並びに、FourEyesを含み、この内の幾つかは、T. Minkaの"An Image Database Browser that Learns from User Interaction," MIT Media Laboratory Perceptual Computing Section, TR#365 (1996)で議論されている。これらのシステムは幾つかの非常によく似通ったものがデータベース内になければならないという名目の下で作業する。この名目の下で、サーチがそのデータベース自体内の要素で開始し、ユーザは所望の画像へ向かって一連の照会例を介してガイドされていく。残念ながら、そうした「ガイド」はユーザが連続してそのサーチをリファイン又は純化しなければならないので、実質的な時間の浪費となる。

階層的グループ分けを先行計算(precompute)するスペース仕切り方式はデータベースのサーチを高速化できるが、そうしたグループ分けは静的であり、新規のビデオがデータベース内に挿入された際に再計算(recomputation)が必要となる。同様に、QBEは原則として伸長性であるが、ビデオ・ショット又はビデオ・ナップショットは一般に多数のオブジェクトを含み、その各々が複雑な多次元特徴ベクトルで記載されている。その複雑性は、形状及び運動の特性を記載する問題によって部分的に生ずる。

サーチ及び検索システムの第2カテゴリー、即ちスケッチに基づく照会システムは、ユーザが描いたスケッチとデータベース内の各画像のエッジ・マップとの相関を計算してビデオ情報を探し求めるものである。Hirata et al. の"Query by Visual Example, Content Based Image Retrieval, Advances in Database Technology-EDBT," 580 Lecture Notes on Computer Science(1992, A Pirotte et al. eds.)に記載されたようなスケッチに基づく照会システムでは、スケッチとデータベース内の各画像のエッジ・マップとの相関を計算している。A Del Bimbo et al. の"Visual Image Retrieval by Elastic Matching of User Sketches,"

19 IEEE Trans. on PAMI, 121-132(1997)では、符合を達成すべくエネルギー汎関数を最小化する技術が記載されている。C. E. Jacobs et al. の"Fast Multiresolution Image Querying,"Proc. of SIGGRAPH, 277-286, Los Angeles(Aug. 1995)では、スケッチの小波シグネチャーとデータベース内の各画像の間の距離を計算している。

幾つかの試みがビデオ・ショットを割り出すべく為されてきたが、何れの試みもビデオ・オブジェクトのダイナミック収集としてビデオ・ショットを表していない。その代わりに、先行技術は、ビデオ・クリップが画像フレームの収集であると仮定することによって、ビデオを単純に割り出すための画像検索アルゴリズムを利用してきた。

特にZhan及びSmoliarによって開発された技術は、QBICで開発された技術と共に、(カラーヒストグラムを用いることによって等の)ビデオ用の画像検索方法を用いる。例えばQBIC方法におけるr-フレーム等の「キー・フレーム」が各ショットが選択される。Zhang及びSmoliarの場合、キー・フレームはクリップから単一フレームを選択することによってビデオ・クリップから抽出される。クリップは、ショットにおける全フレームを平均化して、その平均に最も近いクリップ内のフレームを選択することによって選ばれる。カラーヒストグラム・サーチ等の従来の画像サーチを用いることによって、キー・フレームがビデオを割り出すべく用いられる。

同様に、QBIC投影において、r-フレームは、代表的なフレームとして第1フレーム等の任意フレームを取り込むことによって選択される。ビデオ・クリップが運動を有する場合、モザイクがかけられた表示がそのショットに対する代表的なフレームとして用いられる。QBICはこれらr-フレームに対してそれらの画像検索技術を再度用いてそれらのために、ビデオ・クリップを割り出している。

ビデオ・クリップを割り出すために、インフォメディア投影法はオーディオ・ストリームにスピーチ認識アルゴリズムを用いてビデオのトランスクリプトを作り出している。認識された言葉は、その言葉がはなされたビデオ・フレームに整合させられる。ユーザはキーワード・サーチを行うことによってビデオ・クリッ

ブをサーチできる。しかしながら、重大な失策障壁であると証明されたスピーチ

テキスト変換は、変換アルゴリズムの精度として低く（約20%から30%）、検索品質に対して著しい衝撃である。

上述の先行技術は、自由裁量の所定規準に合致するビデオ情報の特定の片々をサーチして検索できるビデオ・サーチエンジンに基づく効果的なコンテンツに対する増大している必要性を満たし損なっている。この技術は運動ビデオ情報をサーチできないか、或は、そうした情報をパン撮りまたはズーミング等の汎用パラメータに関してのみでサーチできるかの何れかである。同様に先行技術は、空間及び時間の特性（時空間特性）に基づくビデオ情報を検索する技術を説明し損なっている。こうして、上述の現行の技術は、ユーザが定義した照会に応じて、保存されたビデオ情報内に埋め込まれたビデオ・オブジェクトの形状或は運動の特性等、自由裁量の所定規準に合致するビデオ情報の内の特定の片々をサーチして検索できない。

#### 発明の概要

本発明の目的は、ビデオ・サーチエンジンに基づいて真のコンテンツを提供することである。

本発明の更なる目的は、ビデオ情報内に埋め込まれたビデオ・オブジェクトをサーチして検索できるサーチエンジンを提供することである。

本発明の別の目的は、ユーザの調査照会に最もよく符合するオブジェクトのみが検索されることになるように、同定されたビデオ・オブジェクトを濾過するメカニズムを提供することである。

本発明の更に別の目的は、ユーザが定義した照会に応じて、自由裁量の所定規準に合致するビデオ情報の内の特定の片々をサーチして検索できるビデオ・サーチエンジンを提供することである。

また本発明の更なる目的は、運動、色彩（カラー）、並びに、エッジ情報を含むビデオ・オブジェクトの統合された特徴特性に基づくビデオ情報から、ビデオ・オブジェクトを抽出できるサーチエンジンを提供することである。

これらの目的や、以下の詳述される更なる開示を参照して明らかとなる他の目的を満たすために、本発明は、ユーザに、相互作用的なネットワークにわたって、ビデオ・データから成る1つ或はそれ以上の一連のフレームからビデオ・オブジェクト

クトをサーチさせ検索させることを許容するシステムを提供する。このシステムは、有益には、ビデオ・オブジェクト属性の1つ或はそれ以上のデータベースのための記憶装置と、該ビデオ・オブジェクト属性が対応するビデオ・データから成る1つ或はそれ以上連続するフレームのための記憶装置とを含む1つ或はそれ以上のサーバ・コンピュータ、サーバ・コンピュータからビデオ・データから成る1つ或はそれ以上連続するフレームの伝送を許容する通信ネットワーク、並びに、クライアント・コンピュータを含む。クライアント・コンピュータは、運動軌跡情報を含む選択されたビデオ・オブジェクト属性情報を受け取る照会インタフェースと、前記選択されたビデオ・オブジェクト属性情報を受け取り、前記サーバ・コンピュータ内の保存されたビデオ・オブジェクト属性を通信ネットワークによって閲覧又はブラウズして、前記選択されたビデオ・オブジェクト属性に所定閾値内で符合する属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定するブラウザ・インタフェースと、前記サーバ・コンピュータから伝送された、前記決定された1つ或はそれ以上のビデオ・オブジェクトに対応する1つ或はそれ以上のビデオ・データから成る連続フレームを受け取る相互作用ビデオ・プレイヤーとを取容している。

好適な構成において、サーバ・コンピュータに保存されたデータベースは、運動軌跡データベース、時空間データベース、形状データベース、色彩データベース、並びに、質感（テクスチャ）データベースを含む。前記ビデオ・データから成る1つ或はそれ以上の連続フレームは、MPEG-1或はMPEG-2等の圧縮フォーマットでサーバ・コンピュータに保存され得る。

また本システムは、各選択ビデオ・オブジェクト属性をサーバ・コンピュータ内に保存された対応するビデオ・オブジェクト属性と比較して、各ビデオ・オブジェクト属性に対して1つというように、候補ビデオの連続（候補ビデオ・シー

ケンス) から成るリストを作成するようなメカニズムを含むこともできる。同様に、前記候補リストに基づいて、前記選択されたビデオ・オブジェクト属性に所定閾値内で符合する集合的な属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定するメカニズムが、有益的に提供される。またこのシステムは、ビデオ・クリップ内のビデオ・オブジェクト投影のグループに対して、問い合わせ又

は照会における多数のオブジェクト間の空間及び時間の関係(時空間関係)を符合させるメカニズムをも含む。

本発明の第2局面に従えば、少なくとも1つの認識可能な属性を含むビデオ・データから成る連続フレームからビデオ・オブジェクトを抽出する方法が提供される。この方法は、値を決定して、ビデオ・データによって表される少なくとも1つの属性の異なる変形各々にそれら値を割り当てることによって、ビデオ・データから成る現行フレームを量子化して、量子化フレーム情報を生成し、前記属性に基づいて、前記ビデオ・データから成るフレーム上のエッジ検出を実行して、該フレーム内のエッジ点を決定することによって、エッジ情報を生成し、先行するフレームからビデオ情報の1つ或はそれ以上の区分された領域を受け取り、前記受け取られた区分領域を、前記量子化されたフレーム情報及び前記生成されたエッジ情報と比較することによって、前記属性を共有しているビデオ情報の領域を抽出することを要求するものである。

好ましくは、前記抽出するステップは、前記受け取られた領域の内の1つを現行の量子化され、エッジ検出されたフレーム上に投影して、前記領域の任意の移動を時間的に追跡することによって、ビデオ・データから成る現行フレーム内の前記領域を抽出すべくインターフレーム(フレーム相互間の)投影を実行することと、特定の条件下で隣接する抽出領域を現行フレーム内にマージ(merge)又は併合すべくイントラフレーム(フレーム内の)区分を実行することとから成る。この抽出ステップは、隣接領域にイントラフレーム区分した後に残留する現行フレーム内の全てのエッジにラベル表示することをも含むことができ、各ラベル表示されたエッジが現行フレーム内のビデオ・オブジェクトの境界を画成するようにしている。

特別好適な技術において、ビデオ情報の将来フレームをも受け取られて、ビデオ情報から成る現行フレームの光学的フローが、現行フレーム内のビデオ情報から成るブロックと、将来フレーム内のビデオ情報から成るブロックとの間の階層的ブロック突き合わせを実行することによって決定され、そして、ビデオ情報の抽出領域上での運動推定又は運動見積が、光学的フローに基づいて、擬似(アフィン)マトリックスを決定することによって実行される。ビデオ情報の抽出領域は

サイズ及び時間的継続時間に基づくと共に、各領域の擬似モデルに基づいてグループ分けされ得る。

本発明の更に別の局面では、1つ或はそれ以上のビデオ・クリップを含むビデオ・データから成るフレームの連続から、ユーザ入力調査照会に最も符合するビデオ・クリップを探し出す方法であり、所定軌跡内で時間的に移動しているビデオ・オブジェクトを含んでいるようなビデオ・クリップを探し出す方法が提供される。この方法は、少なくとも1つのビデオ・オブジェクト軌跡を規定する調査照会を受け取り、その受け取られた照会と、1つ或はそれ以上の予め規定されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の合計距離を決定し、前記受け取られた照会から最小の合計距離を有する前記規定されたビデオ・オブジェクト軌跡の内の1つ或はそれ以上を選択して、最もよく符合するビデオ・クリップ或は複数のビデオ・クリップを探し出すことを含む。

調査照会及び予め規定されたビデオ・オブジェクト軌跡の双方は正規化され得る。照会正規化ステップは前記受け取られた照会を各正規化ビデオ・クリップへマッピングすることと、前記受け取られマッピングされた照会を前記正規化ビデオ・クリップで規定された各ビデオ・オブジェクト軌跡にまで増倍することとを好ましくは伴う。決定ステップは空間的距離比較か、或は、時空間的距離比較かの何れかによって実現される。

本発明の更なる他の局面において、所定特性をそれぞれが有する1つ或はそれ以上のビデオ・オブジェクトをそれぞれが含む1つ或はそれ以上のビデオ・クリップから、ユーザ入力調査照会に最もよく符合するビデオ・クリップを探し出す方

法が提供される。この方法は、ビデオ・クリップ内の1つ或はそれ以上の異なるビデオ・オブジェクトに対して1つ或はそれ以上の特性を規定する調査照会を受け取り、前記規定された特性の内の少なくとも1つを、所定閾値に符合させるビデオ・オブジェクトを探し出すために、ビデオ・クリップをサーチ又は調査し、前記探し出されたビデオ・オブジェクトから、前記1つ或はそれ以上の異なるビデオ・オブジェクトを含むビデオ・クリップを決定し、前記調査照会で規定された前記1つ或はそれ以上のビデオ・オブジェクトと、前記探し出されたビデオ・オブジェクトとの間の距離を計算することによって、前記決定されたビデオ・ク

リップから最も符合するビデオ・クリップを決定することを含む。これらの特性としては、色彩、質感、運動、サイズ、或は、形状を含み得る。

非常に好適な構成において、ビデオ・クリップは関連されたテキスト情報を含み、調査照会は前記1つ或はそれ以上の異なるビデオ・オブジェクトに対応しているテキスト特性の定義を更に含み、本方法は、そのテキスト特性に符合するテキストを探し出すために前記関連されたテキスト情報を調査するステップを更に含む。次いで、最も符合されたビデオ・クリップは前記決定されたビデオ・クリップ及び前記探し出されたテキストから決定される。

この開示に合体させられると共にその一部を構成する添付図面は、本発明の好適実施例を図示し、本発明の原理を説明する役割を果たす。

#### 図面の簡単な説明

本発明の例示的实施例が、以下、添付図面を参照して詳細に説明される。

図1は、本発明の一局面に従ったビデオ情報をサーチして検索するシステムの概略図であり、

図2は、図1のシステムにおいて有用な照会インタフェースの例示図であり、

図3は、図1のシステムにおいて実行されるビデオ・オブジェクトのサーチ方法の例示図であり、

図4は、本発明の一局面に従ったビデオ情報の連続フレームからビデオ・オブジェクトを抽出するための方法のフローチャートであり、

図5は、図4に示された方法に有用な、領域投影及びインターフレーム・ラベ

ル表示の好適方法のフローチャートであり、

図6は、図4に示された方法に有用なイントラフレーム領域併合の好適方法のフローチャートであり、

図7は、図1のシステムにおいて実行される代替的なビデオ・オブジェクト・サーチ方法の例示図である。

#### 詳細な説明

図1で参照されるように、ユーザが定義した照会に応じて、保存されたビデオ情報内に埋め込まれたビデオ・オブジェクトの形状或は運動の特性等、自由裁量の所定規準に合致するビデオ情報の内の特定の片々をサーチして検索するシステ

ムの例示的实施例が提供されている。このシステム100のアーキテクチャは、サーバ・コンピュータ110、通信ネットワーク120、並びに、クライアント・コンピュータ130の3つの構成要素で広範に構成されている。

サーバ・コンピュータ110は、ビデオ・オブジェクト及びビジュアル特徴に対するメタデータを保存するデータベース111を、オリジナルのオーディオビジュアル情報と抽出されたビデオ・オブジェクト及びビジュアル特徴に関連された任意の関連質感情報とを保存する記憶サブシステム112と共に含む。通信ネットワーク120はインターネット或は広帯域ネットワークに基づかせることができる。よって、図1では1つのコンピュータとして示されているが、サーバ・コンピュータ110はワールドワイドウェブ中に分散された複数のコンピュータであることが可能であり、それら全てが通信ネットワーク120を介してクライアント・コンピュータ130と通信出来る。

クライアント・コンピュータ130は、キーボード131、マウス132、並びに、モニタ133を含み、それらが集まって照会インタフェース及びブラウザ・インタフェースの双方を形成して、ユーザが該コンピュータ130内へ調査照会を入力して、オーディオビジュアル情報を求めてネットワーク100をブラウズ(閲覧)することを許容している。図1に示されていないが、ライトペン及びタッチスクリーン等の他の照会入力装置もこのクライアント・コンピュータ130内へ容易に組み入れることが可能である。モニタ133はネットワーク120を

介してサーバ・コンピュータ110から検索されたビジュアル情報を表示すべく使用されると共に、コンピュータ110のユーザによって入力された調査照会を表示すべく使用される。そうした情報は好ましくは圧縮フォーマット、例えばMP E G-2ビットストリームとして検索されるので、コンピュータ130は適切な市販ハードウェア或はソフトウェア、例えばMP E G-2デコーダを含み、検索された情報を表示可能なフォーマットに圧縮解除又は解凍する。

キーボード131、マウス132等々を用いて、ユーザはコンピュータ130に調査照会を入力して、ビデオ情報のクリップ内に埋め込まれた1つ或はそれ以上のビデオ・オブジェクトの1つ或はそれ以上の調査可能（又は検索可能）な属性を指定する。よって、例えば、もしユーザが特定の軌跡で移動した野球ボール

を含むビデオ・クリップを探し出すことを欲したならば、ユーザはその照会内に含まれるべきオブジェクトの運動134をスケッチして、サイズ、形状、色彩（カラー）、並びに、質感等の追加的に調査可能な属性を選択できる。例示的な照会インタフェースが図2に示されている。

ここで使用されているように、「ビデオ・クリップ」は、限定するものではないが例えば、バットをスイングする野球選手、海を渡って移動するサーフボード、或は、草原地帯を越えて走行する馬等々の、同定可能な属性を有する1つ或はそれ以上のビデオ・オブジェクトを含むビデオ情報から成る連続フレーム（フレーム・シーケンス）を言及するものである。「ビデオ・オブジェクト」は、例えば、質感、色彩、運動、並びに、形状等の対象の1つ或はそれ以上の特徴において均質である複数ピクセルから成る切れ目のない連続した集合である。こうしてビデオ・オブジェクトは、少なくとも1つの特徴に関して一貫性を示す1つ或はそれ以上のビデオ領域によって形成される。例えば歩行する人の（スナップ）ショット（人とはここでの「オブジェクト」）は区分されて、形状、色彩、並びに、質感等の規準に関して異なる隣接領域であるが、それらの運動属性に関しては全てが一貫性を示し得る隣接領域の集まりとなる。

図3を参照すると、調査照会300は、色彩301、質感302、運動303、形状304、形状304、サイズ305、並びに、所望のビデオ・オブジェク

トのパン撮り及びズーム等の汎用パラメータ等の他の属性を含み得る。各属性の相対的な重要性を示す様々な重みも調査照会306に組み入れることができる。調査照会を受けるに及んで、コンピュータ130内のブラウザは、ネットワーク120を介して、サーバ・コンピュータ110のデータベース111内に保存された同様属性を探し求めることになる。サーバ110は幾つかの特徴データベースを含んでおり、即ち、システムが索引を付ける個々別々の特徴各々に対して1つずつのデータベースであり、例えば、色彩データベース311、質感データベース312、運動データベース313、形状データベース314、並びに、サイズ・データベース315である。各データベースは、記憶装置112内に圧縮MPEGビットストリームとして保存されたオリジナルのビデオ情報に関連されている。勿論、他の圧縮フォーマット或は圧縮データが使用可能である。

サーバにおいて、照会された属性の各々は保存された属性と比較されることになり、その詳細な説明は以下に続く。こうして、照会された色彩301は色彩データベース311と突き合わせられ(321)、質感との突き合わせ(322)、運動との突き合わせ(323)、形状との突き合わせ(324)、サイズとの突き合わせ(325)、そして他の任意の属性も同様に行われる。候補ビデオ・ショットのリストがその照会で指定された各オブジェクトに対して生成され、例えば、色彩オブジェクト・リスト331、質感オブジェクト・リスト332、運動オブジェクト・リスト333、形状オブジェクト・リスト334、並びに、サイズ・オブジェクト・リスト335が生成される。サーバ・コンピュータ110において、各リストは予備選択されたランク閾値或は特徴距離閾値と併合され得て、最も可能性のある候補ショットのみが残存するようになる。

次に所定の最小閾値で、各オブジェクトの候補リストが併合されて(350)、単一のビデオ・ショット・リストを形成する。併合プロセスは生成された候補リスト331, 332, 333, 334, 335の各々の比較を伴って、全ての候補リストに現れないビデオ・オブジェクトが排除される。このスクリーニング後に残る候補ビデオ・オブジェクトが、次いで、照会された属性からのそれらの相対的な汎用重み距離に基づいてソート又は分類される。最後に、所定の個々別

々の閾値に基づくと共に、好ましくは照会306で入力されたユーザ定義重みによって変更された汎用閾値が用いられて、オブジェクト・リストが剪定されて、最も符合する候補又は複数候補となる。当方の好適な汎用閾値は0.4である。

併合リスト内のこれらビデオ・ショットの各々に対して、キー・フレームがビデオ・ショット・データベースから動的に抽出されて、ネットワーク120を介してクライアント130側へ戻される。もしユーザがその結果に満足すれば、そのキー・フレームに対応するビデオ・ショットがビデオ・データベースからそのビデオ・ショットを「切り取り」出すことによって該ビデオ・データベースからリアルタイムで抽出され得る。ビデオ・ショットはビデオ・データベースから、その開示がそれを引用することでここに合体されるChang et al. のPCT特許出願第PCT/US97/08266号(1997年5月16日出願)に記載された技術等の圧縮領域におけるビデオ編集方式を用いて抽出される。

当業者であればご理解いただけるように、図3の突き合わせ技術はオブジェクト・レベル或は領域レベルで実行可能である。

図1に関連してここで説明されるシステムに用いられる様々な技術を次に説明する。有意義な調査照会を為すために、クライアント・コンピュータ130は探し出されるべき属性を制限するか或は量子化することができる。こうして、色彩に関して、許容可能な各種色彩の集合がHSV色彩スペースを均一に量子化するが、勿論、特定の色彩が現在のコンピュータでは許容されているように既に量子化されている真の色彩の使用が好ましい。

質感に関しては、周知のMIT質感データベースが各種オブジェクトに対して質感属性を割り当てるべく使用可能である。よってユーザは、そのデータベース内の56の利用可能な質感から調査照会を形成すべく選択しなければならない。勿論、他の質感集合が容易に使用可能である。

ビデオ・オブジェクトの形状は、任意の形状及びサイズの各種卵形を伴う任意の多角形(ポリゴン)であり得る。よってユーザはカーソルの補助で任意の多角形をスケッチでき、円形、楕円、並びに、矩形等の他の周知の形状は予め定義されて、容易に挿入及び操作される。照会インタフェースはこの形状を、その形状

を正確に表している数の集合に翻訳することになる。例えば、円形は中心点及び半径によって表され、楕円形は2つの焦点と距離で表される。

運動に関しては、2つの代替モードが利用可能である。第1として、サーチはビデオ・オブジェクト内のピクセルから成る光学的フローから引き出されるようなビデオ・オブジェクトの感知される運動に基づかせることができる。光学的フローは、汎用運動（即ち、カメラ運動）及びローカル運動（即ち、オブジェクト運動）双方の組み合わせ効果である。例えば、もしカメラが車の運動を追跡していれば、その車はビデオ・シーケンスでは静止しているように見える。

第2として、サーチはビデオ・オブジェクトの「真」の運動に基づかせることができる。この真の運動は、汎用運動が圧縮された後のオブジェクトのローカル運動を言及するものである。移動している車の場合、その車の真の運動はカードライビングの実際の物理的運動である。

支配的な背景シーンの汎用運動は周知の6パラメータの擬似モデルを用いて推

定することができる一方で、階層的なピクセル領域運動推定方法は光学的フローを抽出するために使用される。汎用運動の擬似モデルが用いられて、同一シーンにおける全てのオブジェクトの汎用運動成分を補償する。以下はその6パラメータ・モデルである。

$$dx = a_x \quad (1)$$

$$dy = a_y \quad (2)$$

ここで、 $a_x$ は擬似パラメータであり、 $x$ 、 $y$ は座標であり、 $dx$ 、 $dy$ は各ピクセルでの変位又は光学的フローである。

例えば、ズーム、パン、或は、チルト（傾斜）等の汎用カメラ運動の分類は汎用擬似推定に基づいている。パン撮りの検出のため、汎用運動速度場のヒストグラムは、当業者であればご理解いただけるように、8方向で計算される。もし支配的な数の移動ピクセルを伴う一方向が存在すれば、その方向へのカメラのパンが宣言される。カメラのズーミングは汎用運動速度場の平均的な大きさと上記擬似モデルにおける2つのスケーリング（増倍）・パラメータ（ $a_x$ 及び $a_y$ ）とを試験することによって検出される。十分な運動があり（即ち、平均的な大きさが

所与の閾値以上であり)、 $a_1$ 及び $a_2$ が双方ともに正であり且つ特定の閾値を上回れば、カメラのズームインが宣言される。さもないければ、もし $a_1$ 及び $a_2$ が双方ともに負であり且つ特定値を下回れば、カメラのズームアウトが宣言される。そうした情報は調査照会に含ませることができ、カメラのパン或はズームの有無を示すことになる。

調査は1つ或はそれ以上のビデオ・オブジェクトに関する時間的な情報をも含むことができる。そうした情報はオブジェクトの全体的な継続時間を、相対期間の長い或は短い、或は、絶対期間の秒単位かの何れかで定義し得る。多数のオブジェクト照会の場合、ユーザは、そのシーンにおける様々なオブジェクトの「到着」順序、及び/或は、デッド順序、即ちビデオ・オブジェクトがそのビデオ・クリップから消失する順序を指定することによって全体的なシーンの時間的順序を指定する柔軟性が付与され得る。時間に関する他の有用な属性は倍率又はオブジェクトのサイズが該オブジェクト存続の継続期間にわたって変化する率である。同様に加速度が調査用に適合する属性であり得る。

調査すべくブラウザに対しての実際の照会を形成する前に、様々な属性に重みを付けることができ、その照会においてそれらの相対的重要性を反映させる。特徴の重み付けは全体に動画化されたスケッチに汎用的であり得て、例えば、属性の色彩は全オブジェクトにわたって同一重みを有することができる。システムによって戻されたビデオ・ショットの最終ランクは、ユーザが様々な属性に割り当てた重みによって影響される。

図4を参照しながら、ビデオ・クリップからビデオ・オブジェクトを抽出する技術を以下に説明する。現行フレーム $n+1$ を含んで、圧縮ビデオ情報400から成る連続するフレームで形成されるビデオ・クリップは、図4で例証的に分析されている。

任意のビデオ・オブジェクト抽出に先行して、生ビデオはビデオ・クリップ400等の複数ビデオ・クリップに分けられる。ビデオ・クリップ分割は、上述のChang et al. のPCT特許出願第PCT/US97/08266号に記載されたようなシーン変化検出アルゴリズムによって達成され得る。Chang et al. は、MPEGビッ

トストリームからのDiscrete Cosine Transform係数を用いて統計的測定を計算すべく、突然及び変遷（例えば、ディゾルブ、フェード・イン／アウト、ワイプ）シーン変化の双方を圧縮MPEG-1或はMPEG-2ビットストリームで検出する技術を説明している。次いでこれら測定値は、突然或は変遷シーン変化の試行錯誤的なモデルを認証すべく使用される。

ビデオ・オブジェクトを区分して跡をたどるために、「画像領域」の概念が利用される。画像領域は一貫した特徴、例えば色彩、質感、或は、運動等を伴うピクセルから成る隣接領域であり、車、人、或は、家等の物理的オブジェクトの一部に一般的に対応することになる。ビデオ・オブジェクトは、連続的フレームにおける追跡された画像領域の連続物から構成される。

図4に示される技術は、ビデオ・ショットにおける、エッジ及び運動情報等の静的属性を考慮することによってビデオ・オブジェクトを区分して追跡している。現行フレーム $n$  401は、好ましくは、これから説明される投影及び区分技術430及び運動推定技術440の双方で用いられる。

投影及び区分に先行して、情報は一貫性ある結果を達成するために2つの異なる方法で予め処理される。平行して、現行フレーム $n$ はその情報に対して1つ或はそれ以上の認識可能な属性に基づいて量子化(410)されると共に、エッジ・マップ420を生成すべく使用される。以下に説明されるように当方の好適な具現化では、色彩がその属性として選択され、それは変化する条件下でのその一貫性が故である。しかしながら、情報の他の属性、例えば質感等も同様に、当業者にはご理解いただけるように、投影及び区分プロセスに対する基礎を形成し得ることにある。

図4に示されるように、現行フレーム（即ちフレーム $n$ ）は知覚的で均一な色彩スペース、例えばCIE  $L^*u^*v^*$ スペース内に変換される(411)。RGB等の非均一色彩スペースは、それらスペースにおける距離測定は知覚的な相違と比例しないので、色彩区分に対して適合しない。CIE  $L^*u^*v^*$ 色彩スペースは色彩を1つの輝度チャンネルと2つのクロミナンス・チャンネルとに分割して、それら輝度及びクロミナンスに付与される重みの変化を許容している。これは非常に重要な選

択肢であって、ユーザに所与のビデオ・ショットの特性に従って異なる重みを割り当てる能力を許容している。事実、クロミナンス・チャンネルにより大きい重み、例えば2倍以上の重みを割り当てることは一般的により良好である。

$L^*u^*v^*$ 色彩スペースに変換された情報は、次いで、適応性をもって量子化される(412)。好ましくは、周知のK-Means或はSelf Organization Map clustering algorithms等の量子化技術に基づくクラスタリングが用いられて、 $L^*u^*v^*$ スペース内の実際のビデオ・データから量子化パレットを作り出す。より共通化された固定レベル量子化技術も使用可能である。

適応性ある量子化412の後、非線形メジアン・フィルタリング413が好ましくは使用されて、重要ではない細目や画像内の離れ値力梢り除される一方で、エッジ情報が保存されている。量子化及びメジアン・フィルタリング(又は中央値濾過処理)は、こうして、可能性あるノイズを小さな細目と共に排除することによって画像を簡略化する。

量子化410と同時に、エッジ検出アルゴリズムを用いてフレームnのエッジ・マップが生成される(420)。エッジ・マップは二進法マスクであり、そこでエッジ・ピクセルが1にセットされ且つ非エッジ・ピクセルが0にセットさ

れる。それは周知のCannyエッジ検出アルゴリズムを介して生成され、そのアルゴリズムは画像上で2D(二次元)ガウス予備平滑化を実行して、水平方向及び垂直方向の方向導関数をとる。次いでこれら導関数が用いられて勾配が計算され、ローカル勾配最大値が候補エッジ・ピクセルとして取られる。この出力は2レベル閾値合成プロセスを通じて走らされて、最終エッジ・マップを作り出す。単純なアルゴリズムが利用可能であって、合成プロセスにおけるそれら2つの閾値レベルをその勾配のヒストグラムに基づいて自動的に選択する。

量子化属性情報及びエッジ・マップの双方が投影及び区分ステップ430で利用されて、そこで色彩等の一貫した属性を有する領域が融合される。投影及び区分は、好ましくは、インターフレーム投影431、イントラフレーム投影432、エッジ点ラベル表示432、並びに、簡略化433の4つのサブステップから構成されている。

インターフレーム投影ステップ431は、図4でのフレーム $n-1$ である先行フレームから決定された先行して区分された領域を投影して追跡する。図5で参照される擬似投影ステップ510において、以下に説明されるように、フレーム $n-1$ から存続している領域がそれらの擬似パラメータに従って先ずフレーム $n$ に投影される。もし現行フレームがその連続における第1フレームであれば、このステップは単にスキップされる。次に、変更されたピクセル・ラベル表示プロセス520が適用される。フレーム $n$ における非エッジ・ピクセル毎に、もしそれが投影領域で覆われて、ピクセルの色彩と領域の平均色彩との間の重みが付けられたユークリッド距離、ここでは、 $W_L = 1$ ,  $W_u = 2$ ,  $W_v = 2$ がデフォルト重みであるユークリッド距離が所与の閾値、例えば256以下であれば、ピクセルが旧領域と一貫したラベル表示が為される。もしピクセルがその所与の閾値以下の2つ以上の投影領域で覆われていれば、最も近い距離の領域としてラベル表示される。しかしながら、もし領域その条件を何等満足しなければ、新規のラベルがそのピクセルに割り当てられる。留意することは、エッジ・ピクセルが処理されず、それ故にこの時点でラベル表示されないことである。最後に、接続グラフ530が全てのラベル間、即ち領域間で構築され、即ち、一方の領域におけるピクセルが他方の領域内に隣接するピクセルを有すれば（4接続モード）、それら2つの領域が隣接するものとしてリンクされる。

イントラフレーム投影ステップ432において、先の追跡された新規のラベル表示（領域）がより大きな領域に併合される。図6で参照されるように、反復的な空間的閉じ込めクラスタリング・アルゴリズム610が利用され、そこで所与の閾値よりも小さな色彩距離、好ましくは225を有する2つの隣接する領域が、それら2つの隣接領域間の色彩距離がその閾値よりも大きくなるまで1つの新規領域620に併合される。もし新規領域が2つの隣接領域から生成されると、その平均色彩がそれら2つの旧領域の平均色彩の重みが付けられた平均を取ることによって計算され（630）、それら2つの旧領域のサイズが重量として使用される。次いで領域接続が2つの旧領域の隣接物全てに対して更新される（640）。次いで新規領域は2つの旧領域のラベルから1つのラベルが割り当てられ

(650)、もし双方旧ラベルが先行するフレームからの跡を追っているのであれば、より大きな領域のラベルが選択され、もし一方のラベルが跡を追っていて、他方がそうでなければ跡を追ったラベルが選ばれ、さもなければより大きな領域のラベルが選択される。2つの旧領域が捨てられ(660)、新規領域が何等決定されなくなるまで(670)、このプロセスは繰り返される。

図4に戻ると、エッジ点が領域境界の正確性を確保すべく色彩測定に従ったそれらの隣接領域に割り当てられる(433)。先に議論されたインターフレーム及びイントラフレーム区分プロセスの双方において、非ピクセル・エッジのみが処理されラベル表示される。エッジ・ピクセルは任意の領域に併合されない。これが保証することは、長いエッジによって明らかに分離された領域が空間的に接続されず、それ故に相互に併合され得ないことである。非エッジ・ピクセルの全てのラベル表示の後、同一色彩距離測定に従ってエッジ・ピクセルにそれらの隣接領域に割り当てられる。上述の接続グラフはラベル表示プロセス中に更新され得る。

最後に、簡略化プロセス434が小さな領域、即ち所与数のピクセル以下を伴う領域を削除すべく適用される。その閾値パラメータは画像のフレーム・サイズに準拠している。QCIFサイズ(176×120)画像の場合、好ましいデフォルト値は50である。もし小領域がその隣接領域の内の1つに接近すれば、即ち色

彩距離が色彩閾値以下であれば、その小領域が隣接領域と併合される。さもなければその小領域は排除される。

投影及び区分プロセス430と同時に、現行フレーム $n$ の光学的フローが、引用することでその開示をこの合体させる、M. Bierlingの"Displacement Estimation by Hierarchical BlockMatching,"1001 SPIE Visual Comm. & Image Processing(1988)に記載された技術等の階層ブロック突き合わせ方法を用いて、運動推定ステップ440におけるフレーム $n$ 及び $n+1$ から誘導される。最小平均絶対輝度装置が固定測定ウィンドウ・サイズを用いることによってサーチされるだけである通常のブロック突合せ技術とは異なり、この方法は、稠密な変異ベクトル場(光学的フロー)を推定すべく、階層の異なるレベルで異なるサイズの測

定ウィンドウを用いている。それは相対的に信頼性がある均質な結果を生ずる。3レベル階層を利用することは好ましい。

色彩或は他の属性領域が抽出され、生成されたフレームにおける光学的フローの測定の後、標準線形帰還アルゴリズムが用いられて、各領域に対する擬似運動を推定する(450)。各領域に対して線形帰還が用いられて、擬似運動式を、即ち、その領域内部の稠密運動場に最も近似してフィットする該式中の6つのパラメータを決定する。

擬似運動パラメータは、好ましくは、推定/MPEG圧縮で用いられる共通3ステップ・ブロック突き合わせ技術の拡張である、6次元擬似スペース内の3ステップ領域突き合わせ方法を用いて更にリファイン又は純化される(460)。この周知技術の説明は、Arun N. Netravali et al. の"Digital Pictures: Representation, Compression and Standards, Second Edition"pp.340-344(Plenum Press, New York and London, 1995)に見い出すことができる。各領域に対して、初期擬似モデルが用いられて、最小平均絶対値輝度誤差を伴う領域を投影する新規モデルを探し求める。各次元に沿ったこのサーチはその次元上の初期パラメータの10%として定義される。

擬似運動推定450及び純化460を介して、擬似運動パラメータを伴う均質色彩領域はフレーム $n$ に対して生成される。同様に、これら領域はフレーム $n+1$ の区分プロセスにおいて追跡される。

最後に領域グループ分け470が、過剰区分を回避してより高いレベルのビデオ・オブジェクトを獲得するプロセス中の最終ステップで適用され得る。幾つかの規準が主要な対象領域をグループ分けすべく或は同定すべく採用可能である。

第1として、決定された領域の、ピクセルの平均数であるサイズ、及び、領域が追跡される連続的なフレームの数である継続時間が利用され得て、ノイズ的で重要でない領域を削除する。小さなサイズ及び/或は小さな継続時間の双方を伴う領域が排除され得ることになる。

第2として、同様の運動を具備する隣接領域は1つの移動するオブジェクトにグループ分けされ得る。これが移動するオブジェクトを伴うビデオ・シーケンス

に適用されて、それらオブジェクトを検出する。そうしたグループ分けを実現するために、空間的閉じ込めクラスタリング・プロセスが用いられて、隣接する領域を個々別々のフレームでのそれら擬似運動パラメータに基づいてグループ分けすることができる。次に時間的なサーチ・プロセスが用いられて、異なるフレームでの領域グループを、もしそれら領域グループが少なくとも1つの共通領域を含むのであれば1つのビデオ・オブジェクトとして一体的にリンクすることができる。開始フレームでの各領域グループに対して、そうしたサーチはそのグループ内部の最長の継続時間を伴う領域で始まる。もし領域グループが特定の時間、即ち1秒の $1/3$ 以上で首尾よく追跡されると、新規オブジェクト・ラベル表示がこの領域グループに割り当てられる。最後に、時間的な整合プロセスが適用され得て、ビデオ・オブジェクト内に含まれる領域の一貫性が確保される。もし一領域が僅かな間、例えばそのビデオ・オブジェクト自体の継続時間の10%未満だけ存続すれば、それが領域グループ分けプロセスの過誤として考慮されるべきであり、そのビデオ・オブジェクトから排除される。

図3に関連して先に議論されたように、サーバ・コンピュータ110は、例えば、色彩データベース311、質感データベース312、運動データベース313、形状データベース314、並びに、サイズ・データベース315等の複数の特徴データベースを含み、各データベースはオリジナルのビデオ情報に関連されている。例えば図4を参照して説明した方法によって抽出されたビデオ・オブジェクト等のパース（構文解析）されたビデオ・クリップから抽出された各ビデオ

オブジェクトに対して、付随特徴が有益にはサーバ・コンピュータ110のデータベース内に保存されている。

色彩データベース311に関して、ビデオ・オブジェクト用の代表的な色彩がCIE-LUVスペースで量子化される。量子化は静的プロセスではなく、量子化パレットが各ビデオ・ショットで変化して、色彩変化に依存する。当方の好適構成は代表的な色彩を利用しているが、色彩データベースは単一色彩、平均色彩、色彩ヒストグラム、及び/或は、色彩対をビデオ・オブジェクトに対して含むことも

できる。

質感データベース312に関して、3つの所謂Tamura質感測定値、即ち、粗大、コントラスト、並びに、配向がオブジェクトの質感コンテンツの測定値として計算される。代替的には小波領域質感、質感ヒストグラム、及び/或は、法規フィルタに基づく質感が利用可能であって、データベース312を開発する。

運動データベース313に関して、各ビデオ・オブジェクトの運動は $N-1$ 個のベクトルのリストとして保存され、ビデオ・クリップにおけるフレームの数を $N$ 個とする。各ベクトルは汎用運動補償後の連続フレーム間におけるオブジェクトの質量中心の平均移動である。この情報と共に、当方はオブジェクトの「速度」及びその持続時間の双方を設定することになるビデオ・ショットのフレーム率をも保存する。

形状データベース314に関して、各ビデオ・オブジェクトの形状における主要成分が、E. Saber et al. の "Region-Based affine shape matching for automatic image annotation and query-by-example," 8 Visual Comm. and Image Representative 3-20(1997)で説明されたような十分に理解された固有値分析によって決定される。同時に、領域の第1次及び第2次のモーメントが生成される。2つの他の新規特徴、即ち、正規化面積及びパーセント・面積も計算される。正規化面積は外接円の面積によって分割されたオブジェクトの面積である。もし領域が円によって公正に近似され得れば、そうした近似が為される。例えば、もしオブジェクトの軸線比が0.9以上であり且つ正規化面積も0.9以上であれば、形状は円として分類される。代替的には、幾何学的不変量、各次元の異なる次数のモーメント、多項式近似、スプライン近似、及び/或は、代数学的不変量が利

用可能となる。

最後に、サイズ・データベース315に関して、ピクセルについてのサイズが保存される。

時間に対する空間的関係の評価が編集項目の連続或はオリジナルの間隔グラフとして割り出される。時空データベース等の他のデータベースが使用され得て、

フレーム内のオブジェクト間の間隔関係が間隔グラフ或は2-Dストリップによって割り出される。

次に、調査照会をサーバ・コンピュータ110の特徴データベース111に保存された情報と比較する技術が以下説明される。図3を参照して議論されたように、サーバ110は照会された色彩301、質感322、運動323、形状324、サイズ325、並びに、他の属性をデータベース311, 312, 313, 314, 315等々に保存された情報と突き合わせるタスクを実行して、候補ビデオ・ショット331, 332, 333, 334, 335のリストを生成する。

突き合わせ運動軌跡323に関して、ビデオ・オブジェクトの三次元軌跡が最適に利用される。それはシーケンス  $\{x[i], y[i], t[i]\}$ 、ここで  $i=1, N$  によって表現され、この三次元が2つの空間的次元  $x, y$  と時間的次元  $t$  を含んで、フレーム数に正規化される。フレーム率は真の時間情報を提供する。

クライアント・コンピュータ130では、ユーザがオブジェクト軌跡を、 $x-y$  平面における頂点の連続としてスケッチすることができ、また該オブジェクトのビデオ・クリップ内での継続時間を指定することができる。継続時間又は持続時間はフレーム率と言うことで、長い、中間、並びに、短いの3つのレベルに量子化される。全体的な軌跡は、例えば単位秒当たり30フレーム等のフレーム率に基づいて運動軌跡を均一にサンプリングすることで容易に計算され得る。

本発明の好適な局面に従えば、突き合わせ形跡、空間的モード、並びに、時空モードの内の2つの主要なモードがこれから説明される。空間モードにおいて、運動形跡が  $x-y$  平面上に投影され、その結果として規制された輪郭となる。照会輪郭とデータベース内の各オブジェクトに対する対応する輪郭との間の距離を測定することによって、候補軌跡が決定される。この種の突き合わせは「時間尺度の不変量」を提供し、ユーザがその軌跡を実行すべくオブジェクトによって費

やされる時間が不確かな場合に有用である。

空間-時間モードにおいて、全体的な運動形跡は用いられて、以下の関数に従って距離が計算される。

$$\sum_i ((x_q[i] - x_t[i])^2 + (y_q[i] - y_t[i])^2) \quad (3)$$

ここで、添字  $q$  及び  $t$  は照会及び目標軌跡をそれぞれ言及しており、指数  $i$  はフレーム数にわたって走る。代替的には指数はサブサンプルの集合にわたって走ることになる。

一般に照会オブジェクトの持続時間はデータベース内のそのオブジェクトの持続時間とは異なるので、幾分かの更なる純化が有益であり得る。第1として、それら持続時間が異なる場合、2つの軌跡は2つの持続時間のより短い間だけ符合し得るものであり、即ち、指数  $i$  は照会持続時間及びデータベース持続時間の内の最小値を通じて走ることになる。

代替的には、照会軌跡持続時間及び保存軌跡持続時間は、各々、突き合わせを実行する前に正規持続時間に正規化され得る。例えば、もし各ビデオ・クリップが正規化されて、再生フレーム率が時間に関して所定時間スケール（時間尺度）に増倍されると、調査照会が、その照会をビデオ・クリップにマッピングしてそのマッピングされた照会を照会されたビデオ・クリップで規定されたビデオ・オブジェクト軌跡まで増倍することによって、同一の所定時間スケールに正規化される。

運動の場合のように、照会された色彩 201、質感 222、形状 224、サイズ 225、並びに、他の属性をデータベースに保存された情報と突き合わせるタスクは最適化された比較プロセスを含む。色彩の場合、照会オブジェクトの色彩が、次式(4)に従って、データベース内の候補としての追跡されたオブジェクトの平均色彩と符合させられる。

$$C_d = \sqrt{(L_q - L_t)^2 + 4(U_q - U_t)^2 + 4(V_q - V_t)^2} \quad (4)$$

ここで、 $C_d$  は CIE-LUV スペースにおける重み付けされたユークリッド色彩距離であり、添字  $q$  及び  $t$  は照会及び目標をそれぞれ言及している。

質感の場合、追跡された各オブジェクトに対する3つの Tamura 質感パラメータ

がデータベース 322 に保存されたパラメータと比較される。距離の関数は、各チャンネルに沿っての変動が伴われた、各質感特徴に沿って重みが付けられたユークリッド距離であり、次式(5)の如くに示される。

$$T_d = \sqrt{\frac{(\alpha_q - \alpha_t)^2}{\sigma_\alpha^2} + \frac{(\beta_q - \beta_t)^2}{\delta_\beta^2} + \frac{(\phi_q - \phi_t)^2}{\sigma_\phi^2}} \quad (5)$$

ここで、 $\alpha$ 、 $\beta$ 、並びに、 $\phi$ は粗さ、コントラスト、並びに、配向をそれぞれ言及しており、様々な $\sigma$  ( $\alpha$ ,  $\beta$ ,  $\phi$ )は対応する特徴における変動を言及している。

形状に関して、数式は、次式(6)に示される形状の主要成分のみを単純に含み得る。

$$Sh_d = \left| \frac{\lambda_{2q}}{\lambda_{1q}} - \frac{\lambda_{2t}}{\lambda_{1t}} \right| \quad (6)$$

ここで、及びはオブジェクトの主要軸線に沿っての固有値であり、即ちそれらの比はアスペクト比である。幾何学的な不変量等の他のより複雑なアルゴリズムが使用され得る。

サイズは、次式(7)に示される如くに、距離対面積の比として具現化される。

$$Si_d = 1 - \frac{\min(A_q, A_t)}{\max(A_q, A_t)} \quad (7)$$

ここで、 $A_q$ 、 $t$ は照会及び目標のパーセント面積をそれぞれ言及している。

合計距離は、次式(8)に従って、各関数のダイナミックレンジが[0, 1]内に横たわるに正規化された後、これら距離の重み付けされた合算である。

$$D_q = \sum_{i=1}^n w_i D_i \quad (8)$$

図7を参照すると、埋め込まれたオブジェクト情報及び関連されたオーディオ或はテキスト情報の双方に基づくビデオ・クリップの位置を探し出すべく、サーチ技術に基づき組み合わせられたビデオ及びテキストを以下説明する。この技術は、オブジェクトの運動等のビジュアル・コンテンツや、色彩及び質感等の属性を自然言語の説明能力と共に同時に利用するものである。

色彩701、質感702、運動703、並びに、形状704等のビジュアル属

性の1つ或はそれ以上の入力に加えて、調査照会700に入ると、ユーザはテキスト情報のストリング710を入力することが許容される。こうした情報はキーボード131を通じて、市販の音声認識ソフトウェアに関連するマイクロホンを通じて、或は、コンピュータ・インタフェース技術に対する任意の他の人を通じて、直接的に入力され得る。

ビジュアル情報は、所定閾値に最も符合（マッチ）したビデオ・クリップを生成すべく、図3に関連して議論されたようなビジュアル属性情報の保存されたライブラリ720と突き合わせられる（730）。しかしながら、図7のアーキテクチャは、ビジュアル・ライブラリ720を生成すべく使用された同一のビデオ・クリップに関連された抽出キーワード740でテキスト突き合わせ750を実行することによって、図3に拡張する。テキスト突き合わせ750の結果は、テキストだけに基づいた、1つ或はそれ以上の最もよく符合されたビデオ・クリップである。最後に、ビジュアル突き合わせ730及びテキスト突き合わせ750は組み合わされて（760）、高度の正確性で、オリジナルの調査照会700によって求められたビデオ・クリップを決定する。

MPEG圧縮オーディオビジュアル情報の場合、抽出されたキーワード740のライブラリは手動で注釈が施され得るか、オーディオを転記すべく圧縮ビットストリームから先ずオーディオ情報を抽出してから、キーワード・スポッティング技術でその転記テキストの量を減らすことによって形成され得る。

以上の説明は、本発明に含まれる原理原則の単なる例示である。本発明の他の変更等は当業者であれば明らかあり、本発明の範囲は請求の範囲で詳述されたようにだけ制限されることが意図されている。

【図1】

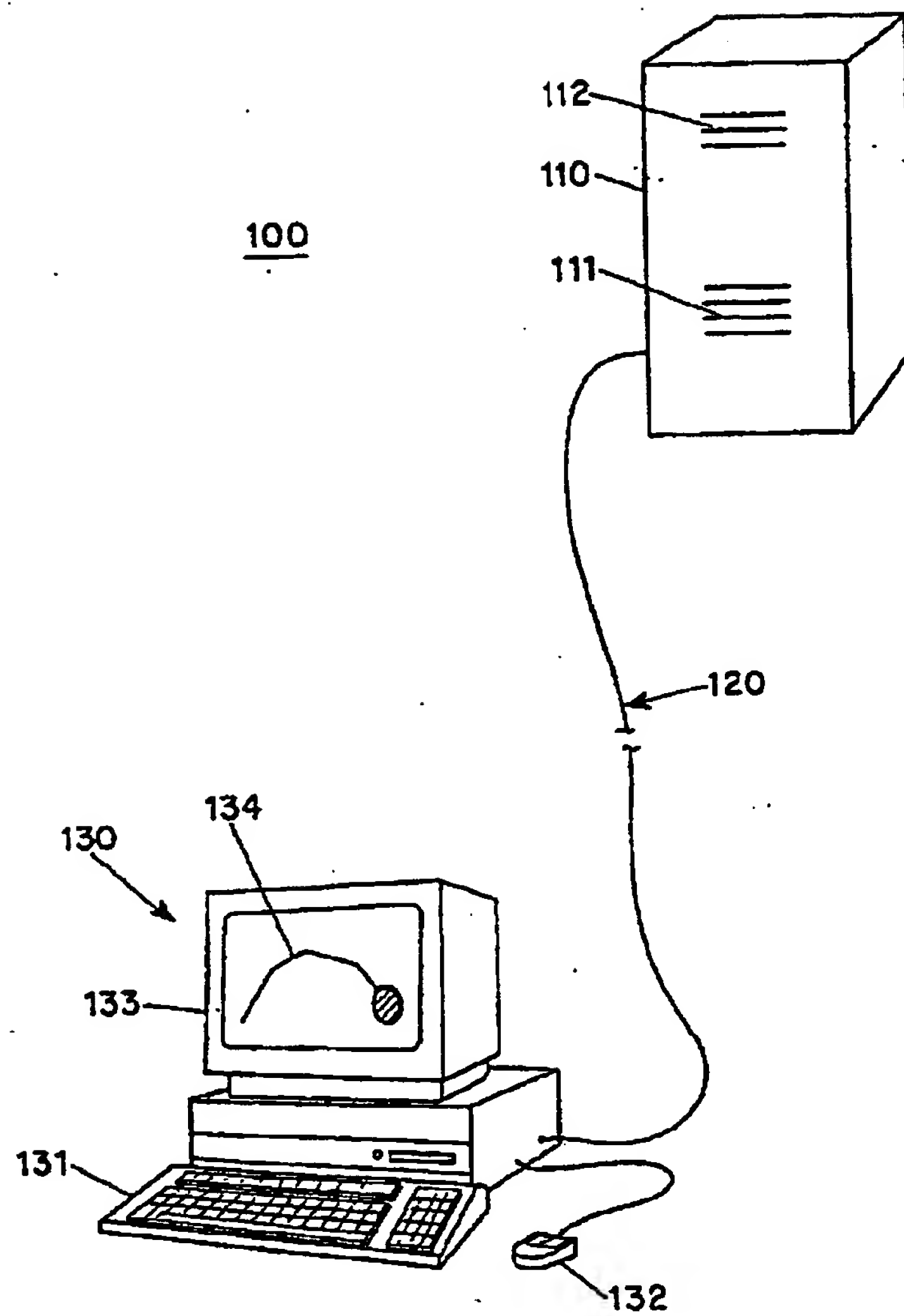


FIG. 1

【図2】

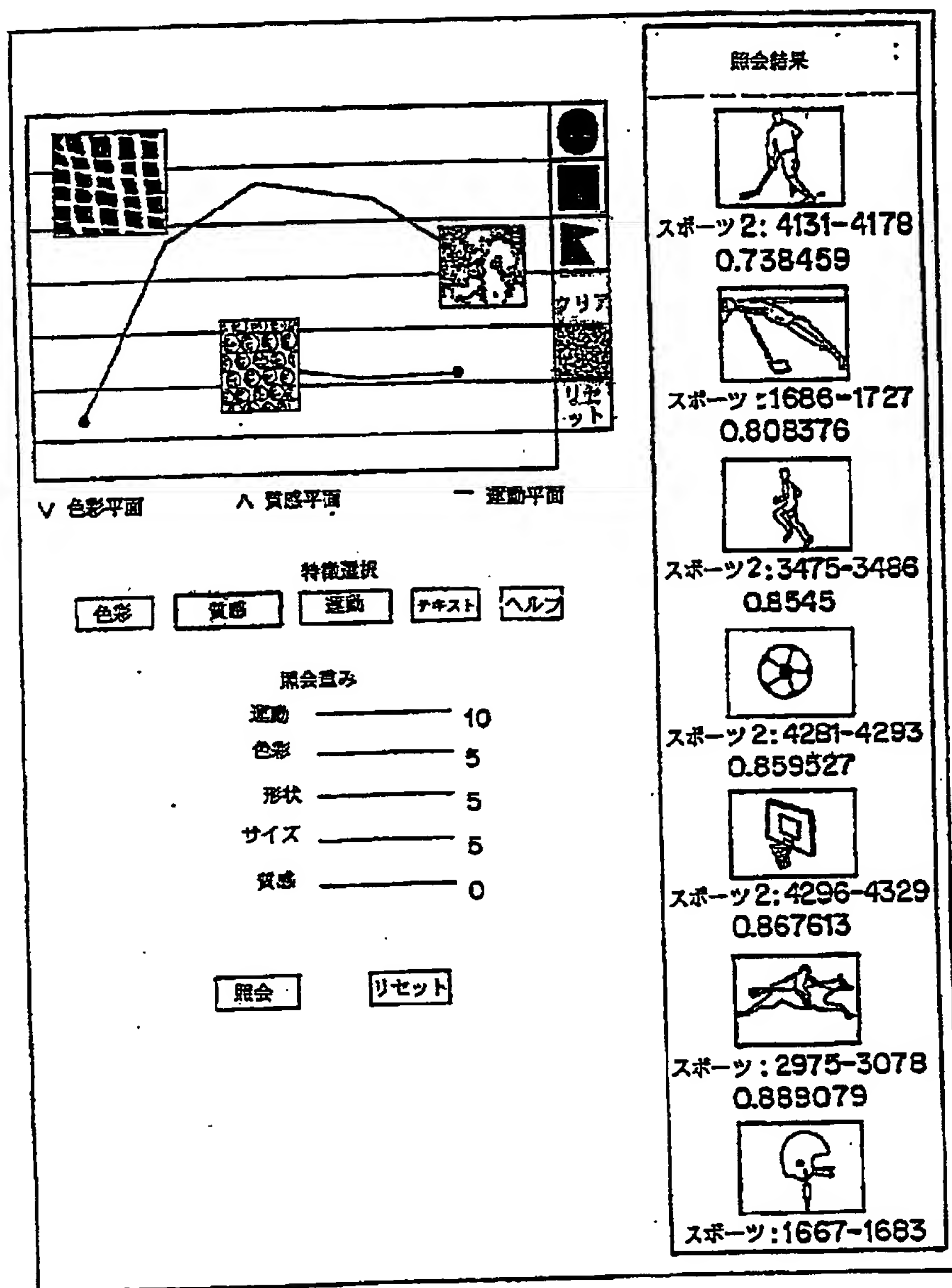


FIG. 2

【図3】

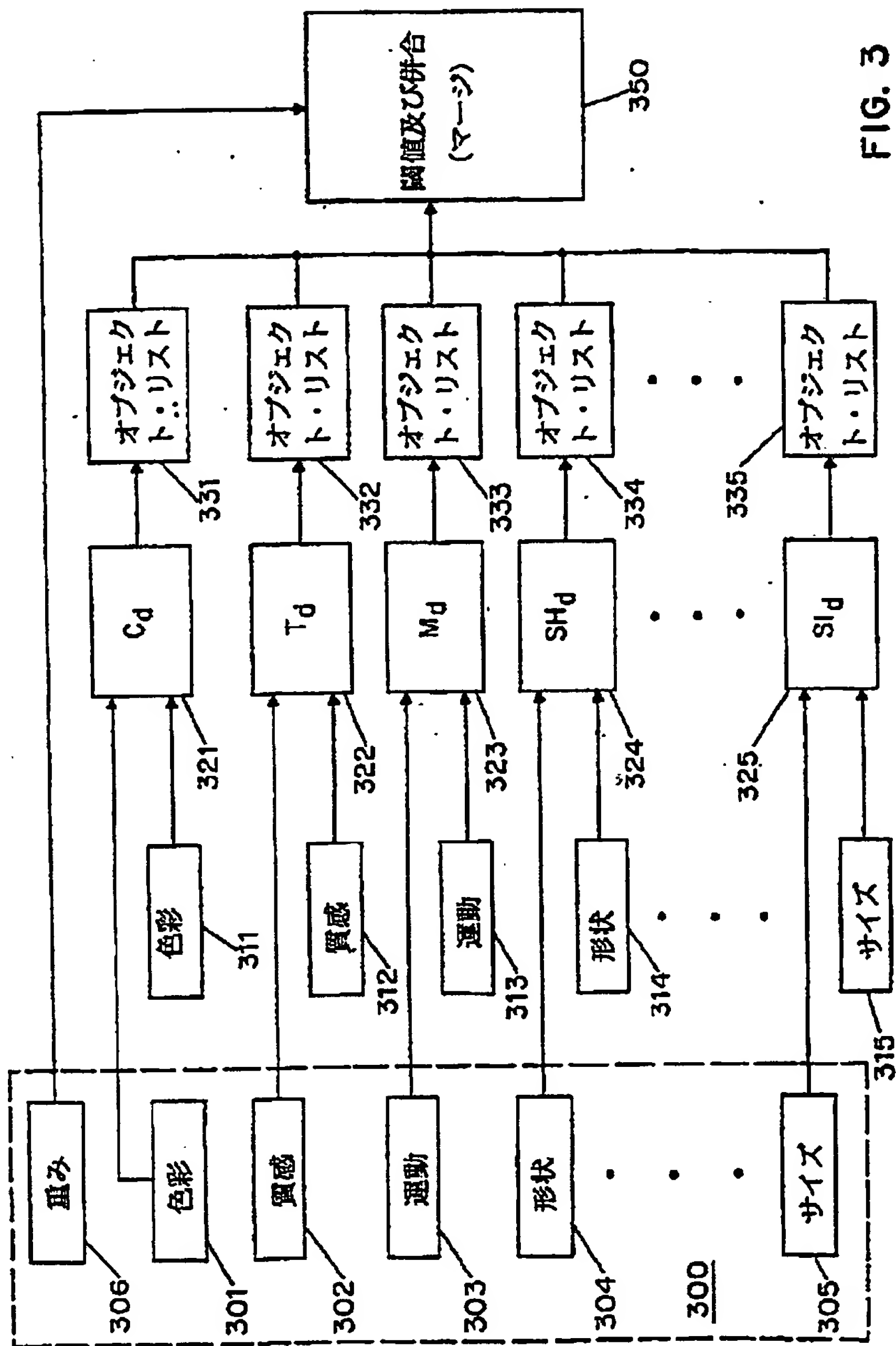


FIG. 3

【図4】

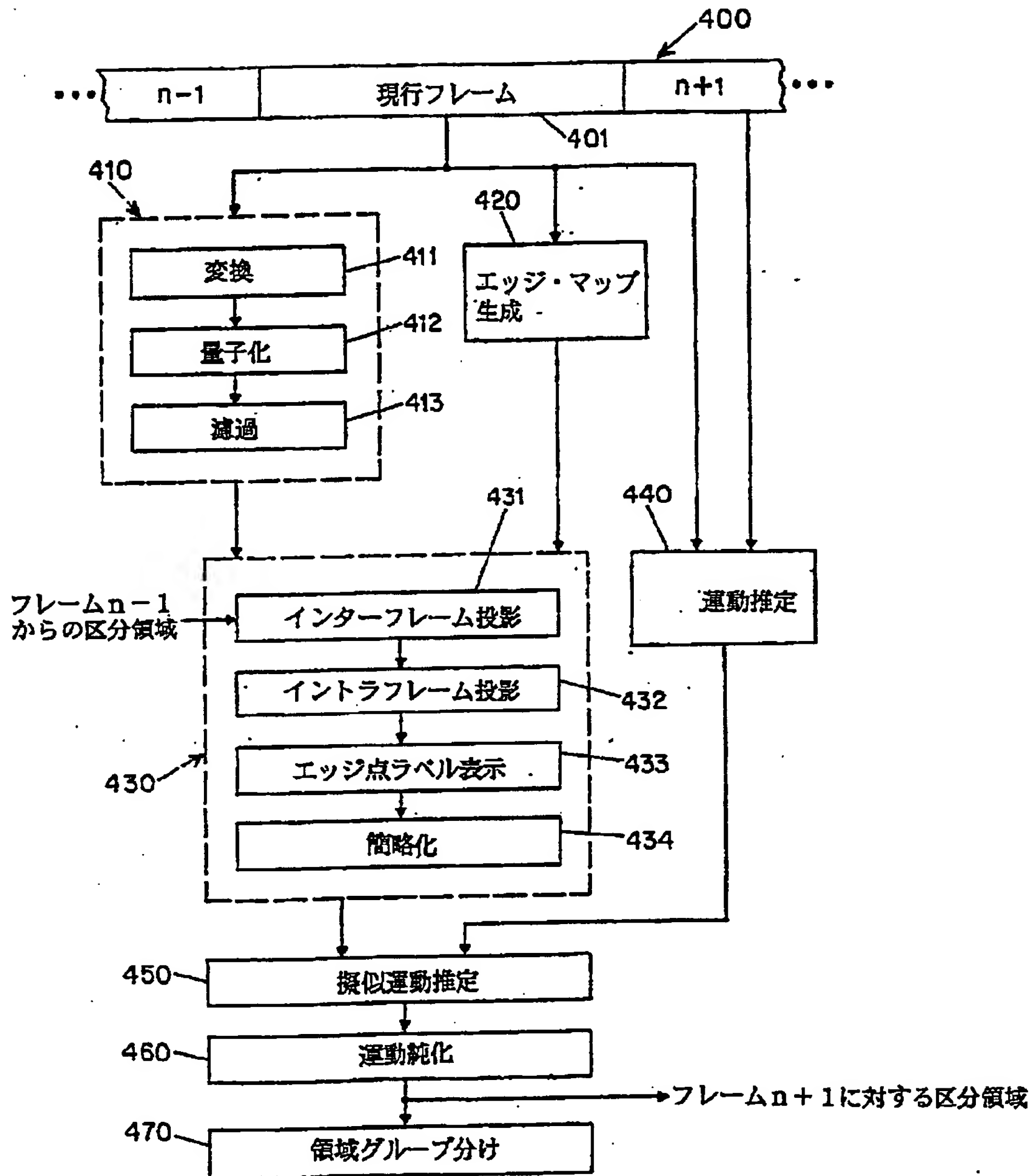


FIG. 4

【図5】

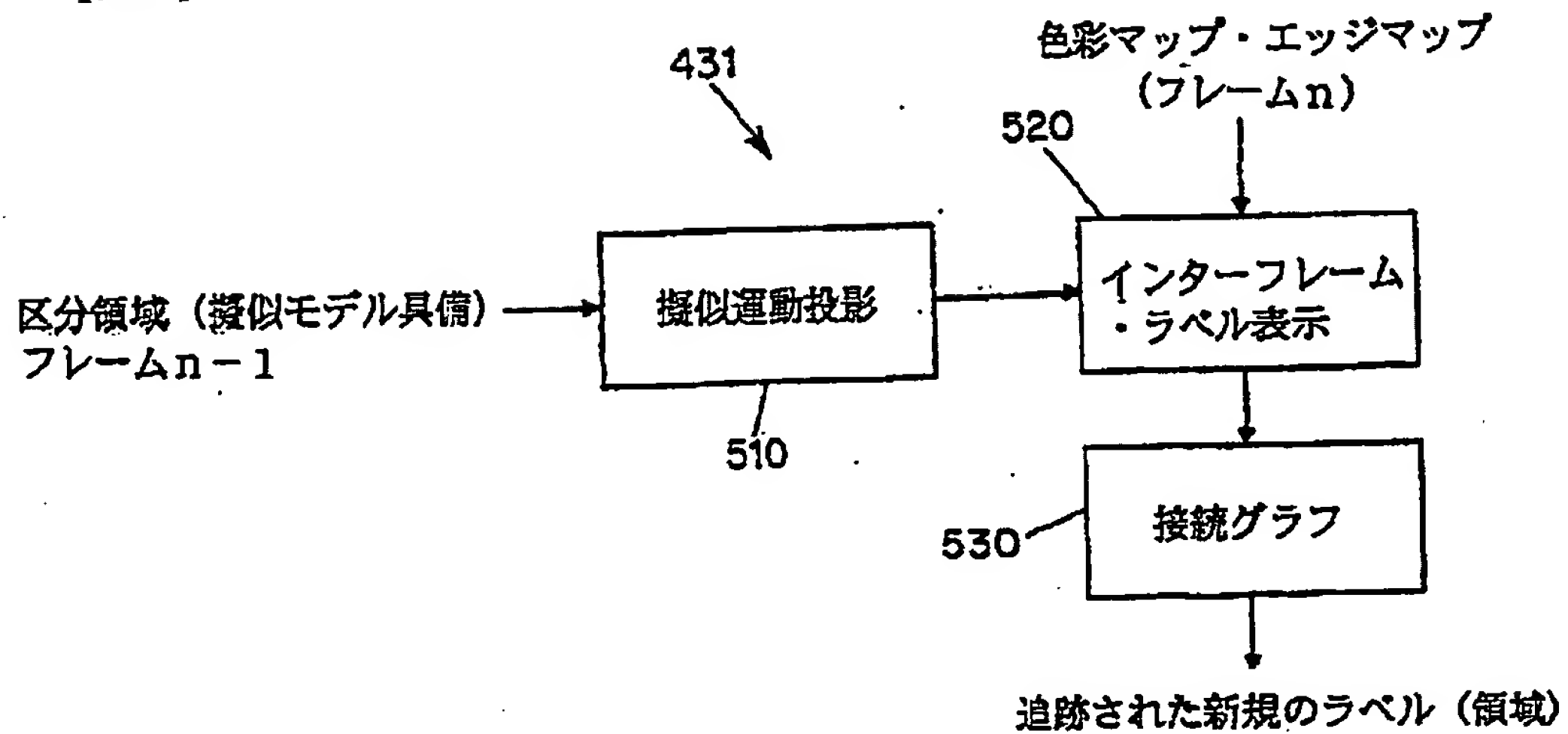


FIG. 5

【図6】

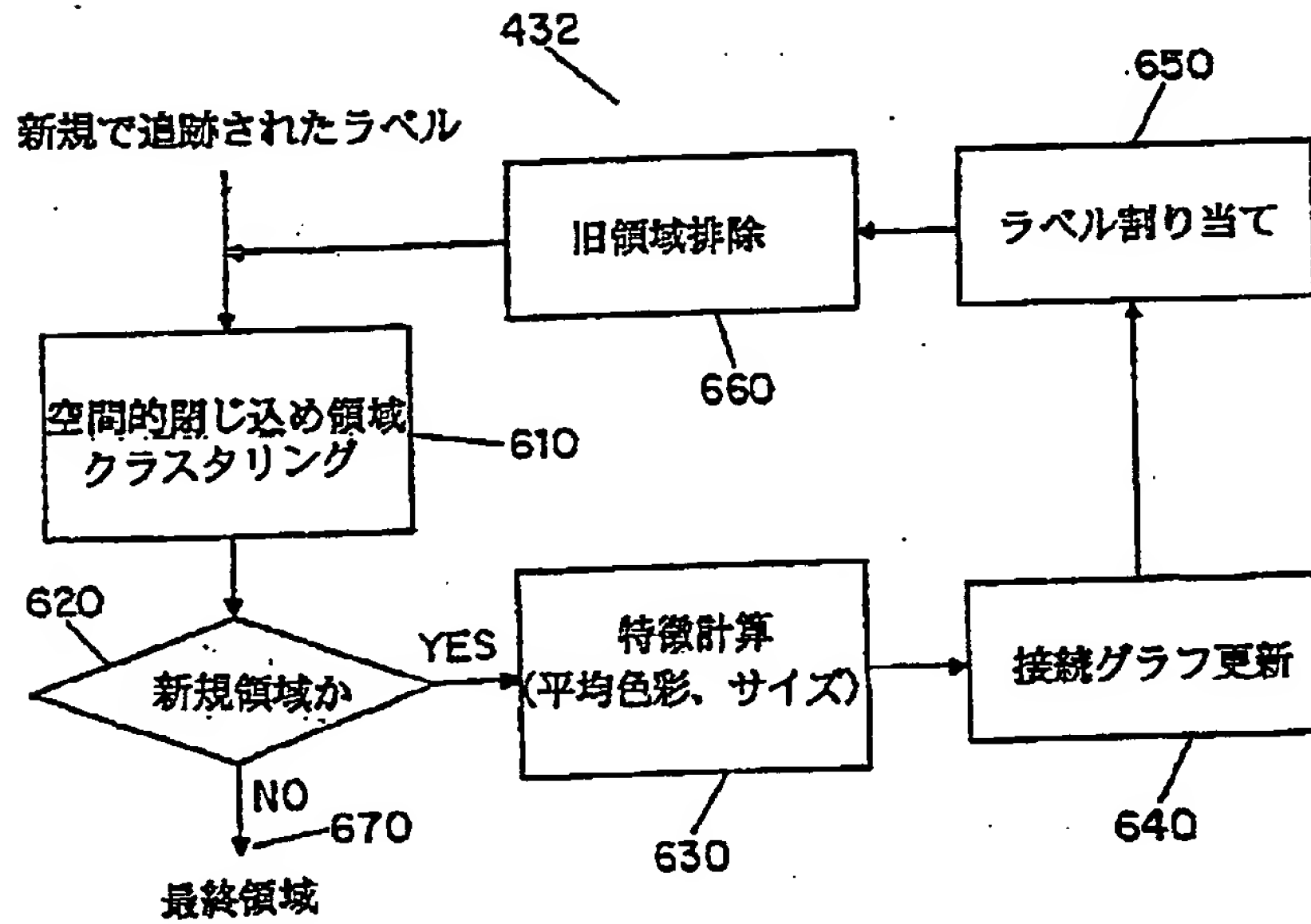


FIG. 6

【図7】

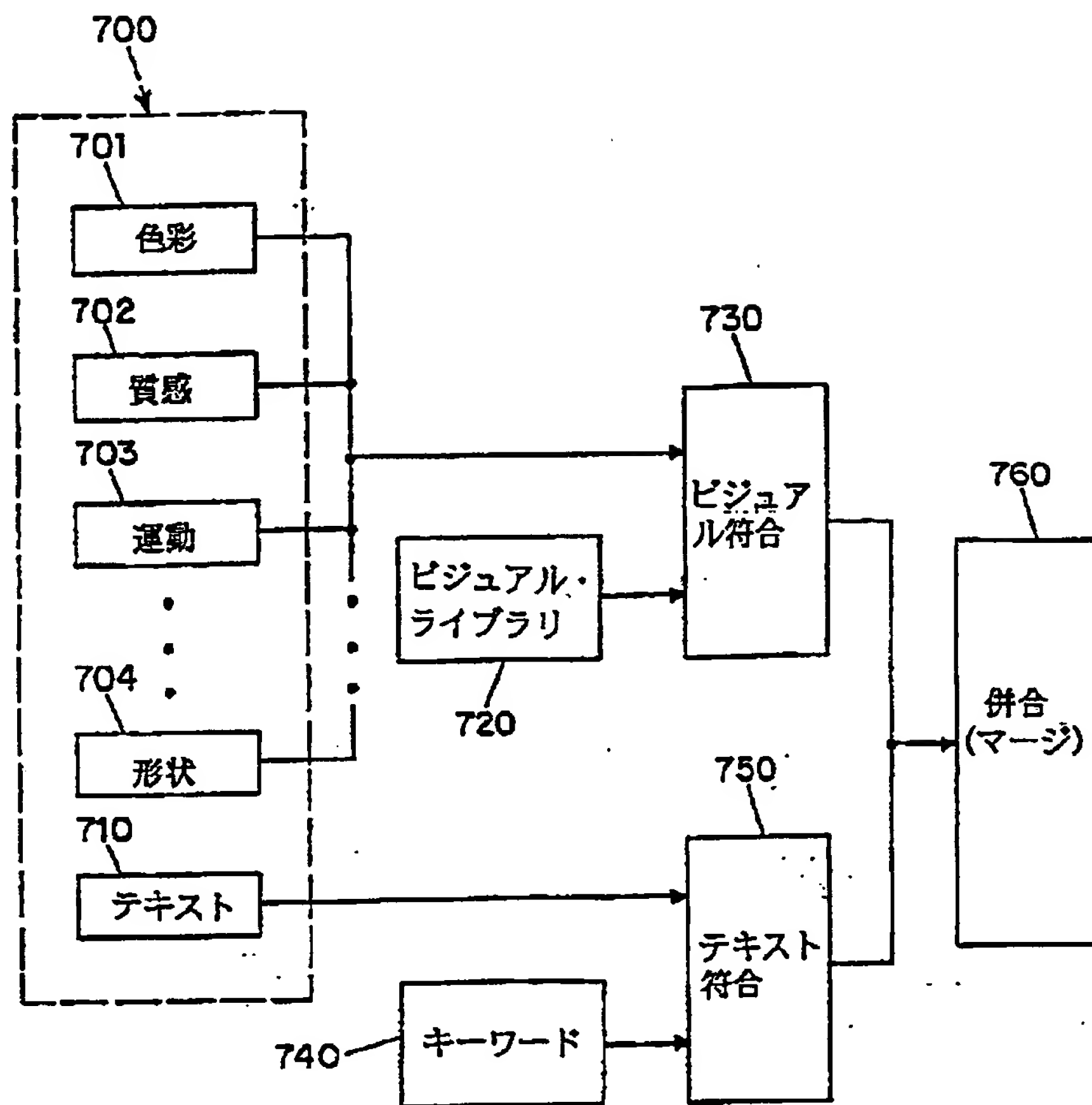


FIG. 7

## 【手続補正書】

【提出日】平成11年11月26日(1999.11.26)

## 【補正内容】

## 請求の範囲

1. 相互作用的なネットワークを介して1つ或はそれ以上のビデオ・クリップから1つ或はそれ以上のビデオ・オブジェクトを視覚的に照会させるとともに探し出させることをユーザに許容するシステムであって、
  - a. 前記1つ或はそれ以上のビデオ・クリップのための記憶装置と、前記ビデオ・クリップに対応するビデオ・オブジェクト属性の1つ或はそれ以上のデータベースのための記憶装置とを備える1つ或はそれ以上のサーバ・コンピュータと、
  - b. 前記1つ或はそれ以上のサーバ・コンピュータに結合された通信ネットワークであり、前記サーバ・コンピュータから前記1つ或はそれ以上のビデオ・クリップの伝送を許容する通信ネットワークと、
  - c. 前記通信ネットワークに結合されたクライアント・コンピュータであり、
    - i. 運動軌跡情報を含むビデオ・オブジェクト属性情報を指定するビジュアル照会インタフェースと、
    - ii. 前記照会インタフェースに結合されて、前記選択されたビデオ・オブジェクト属性情報を受け取るブラウザ・インタフェースであり、前記通信ネットワークによって前記サーバ・コンピュータ内の前記記憶されたビデオ・オブジェクト属性をブラウズして、前記指定されたビデオ・オブジェクト属性に最も符合する属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定するブラウザ・インタフェースと、
    - iii. 前記サーバ・コンピュータから、前記決定された1つ或はそれ以上のビデオ・オブジェクトに対応するビデオ・データから成る1つ或はそれ以上の伝送された連続フレームを受け取る相互作用的なビデオ・プレイヤーとを有するクライアント・コンピュータとを具備するシステム。

2. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、運動軌跡データベースを含む、請求項1に記載のシステム。
3. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、時空間データベースを含む、請求項1に記載のシステム。
4. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、形状データベースを含む、請求項1に記載のシステム。
5. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、色彩データベースを含む、請求項1に記載のシステム。
6. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、質感データベースを含む、請求項1に記載のシステム。
7. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、パン撮りデータベースを含む、請求項1に記載のシステム。
8. 前記サーバ・コンピュータに保存された前記1つ或はそれ以上のデータベースの内の1つが、ズーム撮りデータベースを含む、請求項1に記載のシステム。
9. ビデオ・データから成る前記1つ或はそれ以上の連続フレームが圧縮フォーマットで前記サーバ・コンピュータに保存されている、請求項1に記載のシステム。
10. 前記1つ或はそれ以上の指定されたビデオ・オブジェクト属性の各々を前記サーバ・コンピュータ内に対応して保存されたビデオ・オブジェクト属性と比較して、各ビデオ・オブジェクト属性に1つずつの候補ビデオ・シーケンスのリストを生成する手段を前記サーバ・コンピュータの内の少なくとも1つ或はそれ以上の内部に更に含む、請求項1に記載のシステム。
11. 前記比較手段に結合されて前記候補リストを受け取る手段であり、前記候補リストに基づいて、前記選択されたビデオ・オブジェクト属性に最も符合する集合的な属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定する手段を前記サーバ・コンピュータ内に更に含む、請求項10に記載のシステム。

12. 前記照会されたビデオ・オブジェクト属性情報が、2つ以上のビデオ・オブジェクトに対する属性を含み、前記比較手段が、各ビデオ・オブジェクトに対する前記1つ或はそれ以上の指定されたビデオ・オブジェクト属性の各々を前記サーバ・コンピュータ内の対応する保存されたビデオ・オブジェクト属性と比較して、各ビデオ・オブジェクトに対する各ビデオ・オブジェクト属性に1つずつの、候補ビデオ・シーケンスのリストを生成し、前記決定手段が、各

照会されたビデオ・オブジェクトに対する前記候補リストに基づいて、前記選択されたビデオ・オブジェクト属性に最も符合する集合的属性を有する1つ或はそれ以上のビデオ・オブジェクトを決定している、請求項11に記載のシステム。

13. 少なくとも1つの認識可能な属性を含むビデオ・クリップからビデオ・オブジェクトを抽出する方法であって、

a. ビデオ・データで表示される前記少なくとも1つの属性の異なる変体に値を決定すると共に該値を割り当てることによって、前記ビデオ・データから成る現行フレームを量子化することによって、量子化フレーム情報を生成する工程と、

b. 前記少なくとも1つの属性に基づいて、前記ビデオ・データから成るフレーム上のエッジ検出を実行して、該フレーム内のエッジ点を決定することによって、エッジ情報を生成する工程と、

c. 先行するフレームから1つ或はそれ以上に区分された領域を規定する情報を受け取る工程と、

d. 前記現行フレームから、前記受け取られた区分領域を前記量子化フレーム情報及び前記生成エッジ情報と比較することによって、前記少なくとも1つの属性を共有するビデオ情報から成る領域を抽出する工程との諸工程を含む方法。

14. 前記属性が色彩であり、量子化工程が、前記現行フレームを均一色彩スペース情報に変換し、該色彩スペース情報を適応的に量子化し、前記パレットを濾過してそこからノイズを排除することを含む、請求項13に記載の方法。

15. 前記適応的な量子化工程が、クラスタリング・アルゴリズムでの量子化を含む、請求項14の方法。
16. 前記エッジ検出工程が、前記現行フレームにCannyエッジ検出を適用して、前記エッジ情報をエッジ・マップとして生成することを含む、請求項13に記載の方法。
17. 前記抽出工程が、
- a. 前記受取領域の内の1つを現行の量子化され且つエッジ検出されたフレームに投影することによって、ビデオ・データから成る前記現行フレーム内の領域を抽出すべくインターフレーム投影を実行して、前記領域の任意の移動を時間的に追跡することと、
  - b. 前記現行フレーム内の隣接する抽出領域を併合すべくイントラフレームを実行することと、
- を含む、請求項13に記載の方法。
18. 前記属性が色彩であり、前記インターフレーム投影工程が、
- a. 前記先行フレームからの前記受取領域を前記現行フレーム内へ投影して、領域を時間的に追跡する工程と、
  - b. 前記受取領域と一貫性を有するか或は新規領域として、前記現行フレーム内における各非エッジ・ピクセルをラベル表示する工程と、
  - c. 隣接する領域をリンクすべく、前記ラベルから接続グラフを生成する工程と、
- の諸工程を含むことから成る、請求項17に記載の方法。
19. 前記イントラフレーム区分工程が、
- a. 所定閾値よりも小さい色彩距離を有する全ての隣接領域を新規領域として併合する工程と、
  - b. 前記新規領域に対する平均色彩を決定する工程と、
  - c. 前記接続グラフを更新する工程と、
  - d. 前記新規領域に、前記併合領域に先行して割り当てられたラベルから新規ラベルを割り当てる工程と、

e. 前記併合領域を排除する工程と、

の諸工程を含むことから成る、請求項18に記載の方法。

20. 前記抽出工程が、隣接領域に為す前記イントラフレーム区分の後に残存する前記現行フレーム内の全てのエッジをラベル表示して、それらの各ラベル表示されたエッジが前記現行フレーム内のビデオ・オブジェクトの境界を画成するように為す工程を更に含むことから成る、請求項17に記載の方法。

21. 前記抽出工程が、所定閾値以下のサイズを有する領域を削除することによって、前記抽出領域を簡略化する工程を更に含むことから成る、請求項20に

記載の方法。

22. e. ビデオ情報から成る将来フレームを受け取る工程と、

f. 前記現行フレーム内のビデオ情報から成るブロックと、前記将来フレーム内のビデオ情報から成るブロックとの間の階層的なブロック突き合わせを実行することによってビデオ情報から成る前記現行フレームの光学的フローを決定する工程と、

g. 前記光学的フローに基づいて、ビデオ情報から成る前記抽出領域上の運動推定を実行する工程と、

の諸工程を更に含む、請求項13に記載の方法。

23. 前記現行フレーム内の決定領域をサイズ及び持続時間でグループ分けする工程を更に含む、請求項22に記載の方法。

24. 前記現行フレーム内の決定領域をその内部の移動オブジェクトを決定することでグループ分けする工程を更に含む、請求項22に記載の方法。

25. それぞれが所定軌跡で時間的に移動する1つ或はそれ以上のビデオ・オブジェクトを含んでいる1つ或はそれ以上のビデオ・クリップから、ユーザ入力ビジュアル調査照会に最も符合するビデオ・クリップを探し出す方法であって、

a. 少なくとも1つのビデオ・オブジェクト軌跡を画成するビジュアル調査照会を受け取る工程と、

b. 前記受取照会と、1つ或はそれ以上の予め画成されたビデオ・オブジェ

クト軌跡の内の少なくとも一部との間の合計距離を決定する工程と、

c. 前記受取照会から最小距離を有する前記画成されたビデオ・オブジェクト軌跡の内の1つ或はそれ以上を選択することによって、前記最も符合したビデオ・クリップ或は複数のビデオ・クリップを探し出す工程と、

の諸工程を含む方法。

26. 前記保存されたビデオ・クリップが正規化されて、再生フレーム率が所定時間スケールまで増倍されるように為され、前記受取照会を各正規化ビデオ・クリップへマッピングして、そのマッピングされた受取照会を前記正規化ビデオ・クリップによって規定された各ビデオ・オブジェクトまで増倍することによって前記受取調査照会を正規化する工程を更に含み、前記決定工程が

前記正規化された受取照会と前記正規化されたビデオ・オブジェクト軌跡との間の合計距離を決定してることから成る、請求項25に記載の方法。

27. 前記決定工程が、前記受取ビデオ・オブジェクト軌跡と前記1つ或はそれ以上の予め規定されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の空間的距離の比較を為すことを含むことから成る、請求項25に記載の方法。

28. 前記決定工程が、前記受取ビデオ・オブジェクト軌跡と前記1つ或はそれ以上の予め規定されたビデオ・オブジェクト軌跡の内の少なくとも一部との間の時空間的距離の比較を為すことを含むことから成る、請求項25に記載の方法。

29. それぞれが所定属性を有する1つ或はそれ以上のビデオ・オブジェクトを含んでいる1つ或はそれ以上のビデオ・クリップから、ユーザ入力ビジュアル調査照会に最も符合するビデオ・クリップを探し出す方法であって、

a. 前記ビデオ・クリップ内の1つ或はそれ以上の異なるビデオ・オブジェクトに対する1つ或はそれ以上の属性を規定するビジュアル調査照会を受け取る工程と、

b. 前記ビデオ・クリップを調査して、前記規定された属性の内の少なくとも1つを所定閾値に符合させる1つ或はそれ以上のビデオ・オブジェクトを探し出す工程と、

c. 前記探し出されたビデオ・オブジェクトから、前記1つ或はそれ以上の異なるビデオ・オブジェクトを含む1つ或はそれ以上のビデオ・クリップを決定する工程と、

d. 前記調査照会によって規定された前記1つ或はそれ以上のビデオ・オブジェクトと前記探し出されたビデオ・オブジェクトとの間の距離を計算することによって、前記決定されたビデオ・クリップから最も符合するビデオ・クリップを決定する工程と、  
の諸工程を含む方法。

30. 前記1つ或はそれ以上の属性が色彩を含み、前記符合工程が、前記照会されたビデオ・オブジェクトの各々に対する平均色彩を決定して、該平均色彩をデータベースに保存された色彩情報と比較することを含むことから成る、請求項29に記載の方法。

31. 前記1つ或はそれ以上の属性が質感を含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対する粗さ、コントラスト、並びに、配向を決定して、該粗さ、コントラスト、並びに、配向をデータベース内に保存された粗さ、コントラスト、並びに、配向の情報と比較することを含むことから成る、請求項29に記載の方法。

32. 前記1つ或はそれ以上の属性が形状を含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対する主要軸線に沿っての固有値を決定して、該固有値をデータベース内に保存された形状情報と比較することを含むことから成る、請求項29に記載の方法。

33. 前記1つ或はそれ以上の属性がサイズを含み、前記符合工程が、前記照会されたビデオ・オブジェクト各々に対するパーセンテージ面積を決定して、該面積をデータベース内に保存された面積情報と比較することを含むことから成る、請求項29に記載の方法。

34. 前記ビデオ・クリップが関連されたテキスト情報を含み、前記参照照会が前記1つ或はそれ以上の異なるビデオ・オブジェクトに対応しているテキスト特性の定義を更に含み、前記関連されたテキスト情報を調査して前記テキスト

特性に最も符合するテキストを探し出す工程を更に含む、請求項29に記載の方法。

35. 前記最も符合するビデオ・クリップが前記決定ビデオ・クリップ及び前記探し出されたテキストから決定されることから成る、請求項30に記載の方法

。

## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/09124

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06F 17/30, 17/00, 15/00; G06K 9/52

US CL : 707/102, 104; 382/159; 345/325, 328, 440

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/102, 104; 382/159, 218, 203; 345/325, 328, 440

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, INSPEC, WIDS, JAPRO, PATOSWO, PATOSEP

search terms: content-based, video clip, search or query, motion

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	CHANG, S.-F. et al. VideoQ: An Automated Content-Based Video Search System Using Visual Cues. Proceedings ACM Multimedia 97, Seattle, WA, 9-13 Nov. 1997, pages 313-324, see entire document.	1-35
Y,P	CHANG, S.-F. Content-Based Indexing and Retrieval of Visual Information. IEEE Signal Processing Magazine, July 1997, Vol 14, No. 4, pages 45-48.	1-12, 25-35
X,P — Y,P	US 5,734,893 A (LI ET AL) 31 March 1998, see entire document, especially column 3, line 55 to column 6 line 45.	1-6, 9-12, 25, 29-30 — 7-8, 13-24, 26-28, 31-35

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents	T Inter-document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A document defining the general state of the art which is not considered to be of particular relevance	X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E earlier document published on or after the international filing date	Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z document member of the same patent family
*O document referring to an oral disclosure, use, exhibition or other means	
*P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

12 JUNE 1998

Date of mailing of the international search report

08 OCT 1998

 Name and mailing address of the ISA/US  
 Commissioner of Patents and Trademarks  
 Box PCT  
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

David K Moore

Telephone No. (703) 308-7452

## INTERNATIONAL SEARCH REPORT

 International application No.  
 PCT/US98/09124

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 5,655,117 A (GOLDBERG ET AL) 05 August 1997, col. 2, line 5 to col. 4, line 29.	1-12, 25, 29
X	LI, W. et al. VISION: A Digital Video Library. Proceedings of the 1st ACM International Conference on Digital Libraries, Bethesda, MD, 20-23 March 1996. pages 19-27, see entire document.	29, 34-35
Y		1-12, 25-28, 30-33
X	GONG Y. et al. A Generic Video Parsing System with a Scene Description Language (SDL). Real-Time Imaging, February 1996, vol 2, No. 1, pages 45-59, see entire document,	1-12, 25, 29-30, 33
Y		31-32
Y	US 5,606,655 A (ARMAN ET AL) 25 February 1997, see entire document, especially col. 4, lines 8-37, col. 9, lines 6-48	13-14, 22-24

## フロントページの続き

- (72)発明者 チェン ウィリアム  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク ウェスト ワンハンドレ  
ッドトゥエルフス ストリート 423 ア  
partment 34エイ
- (72)発明者 メン ホラス ジェイ  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク ウェスト ワンハンドレ  
ッドナインティーンズ ストリート 435  
アパートメント 9エル
- (72)発明者 サンダラム ハリ  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク ウェスト ワンハンドレ  
ッドトゥエンティース ストリート 434  
アパートメント 9ディー
- (72)発明者 ツォン ディ  
アメリカ合衆国 ニューヨーク州 10027  
ニューヨーク ウェスト ワンハンドレ  
ッドナインティーンズ ストリート 400  
アパートメント 15ディー

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**